Psychometrika

A JOURNAL DEVOTED TO THE DEVELOPMENT OF PSYCHOLOGY AS A QUANTITATIVE RATIONAL SCIENCE

THE PSYCHOMETRIC SOCIETY - ORGANIZED IN 1935

VOLUME 20
NUMBER 3
SEPTEMBER
1 9 5 5

PSYCHOMETRIKA, the official journal of the Psychometric Society, is devoted to the development of psychology as a quantitative rational science. Issued four times a year, on March 15, June 15, September 15, and December 15.

SEPTEMBER, 1955, VOLUME 20, NUMBER 3

Published by the Psychometric Society at 1407 Sherwood Avenue, Richmond 5, Virginia. Entered as second class matter at the Post Office of Richmond, Virginia. Editorial Office, Department of Psychology, The University of North Carolina, Chapel Hill, North Carolina.

Subscription Price: The regular subscription rate is $14.00 per volume. The subscriber receives each issue as it comes out, and, upon request, a second complete set for binding at the end of the year. All annual subscriptions start with the March issue and cover the calendar year. All back issues but two are available. Back issues are $14.00 per volume (one set only) or $3.50 per issue, with a 20 per cent discount to Psychometric Society Members. Members of the Psychometric Society pay annual dues of $7.00, of which $6.30 is in payment of a subscription to *Psychometrika*. Student members of the Psychometric Society pay annual dues of $4.00, of which $3.60 is in payment for the journal.

Application for membership and student membership in the Psychometric Society, together with a check for dues for the calendar year in which application is made, should be sent to

MELVIN D. DAVIDOFF
Test Development Section, U.S. Civil Service Commission, Washington 25, D. C.

Payments: All bills and orders are payable in advance.

Checks covering membership dues should be made payable to the *Psychometric Society*.

Checks covering regular subscriptions to *Psychometrika* (for non-members of the Psychometric Society) and back issue orders should be made payable to the *Psychometric Corporation*. All checks, notices of change of address, and business communications should be sent to

DAVID R. SAUNDERS, Treasurer, Psychometric Society and Psychometric Corporation
Educational Testing Service
P.O. Box 592
Princeton, New Jersey

Copies not received because of failure to report change of address will be replaced, upon request, and will be billed at the appropriate back issue rate. Allow at least six weeks for notification of change of address to take effect, sending notice to the Treasurer of the Psychometric Corporation.

Articles on the following subjects are published in *Psychometrika:*

(1) the development of quantitative rationale for the solution of psychological problems;
(2) general theoretical articles on quantitative methodology in the social and biological sciences;
(3) new mathematical and statistical techniques for the evaluation of psychological data;
(4) aids in the application of statistical techniques, such as nomographs, tables, work-sheet layouts, forms, and apparatus;
(5) critiques or reviews of significant studies involving the use of quantitative techniques.

The emphasis is to be placed on articles of type (1), in so far as articles of this type are available.

(*Continued on the back inside cover page*)

Psychometrika

CONTENTS

VOLUME TWENTY SEPTEMBER 1955 NUMBER 3

Sir Godfrey Thomson

Sir Godfrey Thomson

It was at the International Congress of Psychology, 1923, that I first met Godfrey Thomson. We were in the same symposium on the nature of intelligence. In correspondence and in personal conferences I have found him always friendly and intellectually generous even when we did not agree in our psychological interpretations. I always read his criticisms with interest and respect. An outstanding characteristic was that he never falsified a problem in order to win an argument—a trait that was not shared by some of his adversaries in the controversies of mental measurement.

Godfrey Thomson was born in Carlisle, England, on March 27, 1881. He was educated at Rutherford College, Armstrong College (now King's College), the University of Durham, and the University of Strasbourg. At Armstrong College he was Open Exhibitioner, Junior Pemberton Scholar, and Charles Mather Scholar. Later he was appointed Pemberton Fellow of the University of Durham, where he obtained the M.Sc. degree in mathematics and physics. Following this he attended the University of Strasbourg in Germany and was awarded the Ph.D., *summa cum laude*, in 1906.

At this point his interest turned from the physical sciences to psychology and he returned to the University of Durham for postgraduate study in that subject. After receiving the D.Sc. degree in Psychology in 1913 he accepted the position of Lecturer in Education at Armstrong College. In 1920 he became Professor and Head of the Department of Education; he held this position until 1925. During this period he visited the United States as Visiting Professor of Education at Columbia University, 1923–24. A second visit to this country was in 1933 when he was a lecturer in the Yale Summer School.

From 1925 until his retirement in 1951 he held the joint post of Professor of Education at the University of Edinburgh and Director of Studies, Edinburgh Provincial Committee for the Training of Teachers. In 1939 the University of Durham awarded him an Honorary D.C.L. Later he was awarded the Order of Polonia Restituta (third class) by the Government of Poland in exile, and in 1949 he was knighted. Sir Godfrey Thomson died in Edinburgh on February 9, 1955, at the age of 73.

Godfrey Thomson was a fellow of the Royal Society of Edinburgh, of the Eugenics Society, and of the British Psychological Society, of which he was president, 1945–46. He was an Honorary Fellow of the Educational Institute of Scotland, and of the Swedish Psychological Society. He was a member of the British Association for the Advancement of Science, the National Institute of Industrial Psychology, the International Statistical Institute, and of a large number of boards and foundations.

Sir Godfrey Thomson had many connections with scientific societies in the United States: Foreign Honorary Member of the American Academy of

Arts and Sciences, Foreign Associate of the United States National Academy of Sciences, Fellow of the American Association for the Advancement of Science, member of the American Institute of Mathematical Statistics, and member of the Psychometric Society.

He devised tests of intelligence and achievement which were well-known and widely used in the British Isles and throughout the Commonwealth. With the profits from the sale of these tests he founded scholarships and endowed the Godfrey Thomson Lectureship in Educational Research in Edinburgh University.

Sir Godfrey Thomson's work in mental measurement can be divided into three successive periods. First he was interested in psychophysical problems, beginning in 1911. His work was published in *Essentials of Mental Measurement* by Brown and Thomson and in a number of papers. The second period represents his work on the social and geographical distribution of intelligence and the influence of differential birth rate. A third period was devoted to the factorial analysis of human ability, a field which interested him the most. His work in this field is represented by his well-known book *The Factorial Analysis of Human Ability*, which has appeared in several editions. He described his main objective as an attempt "to bring mathematical exactitude into psychological experiment and theorizing."

Psychometric Laboratory *L. L. Thurstone*
University of North Carolina

PSYCHOMETRIKA—VOL. 20, NO. 3
SEPTEMBER, 1955

A GENERALIZED SIMPLEX FOR FACTOR ANALYSIS*

LOUIS GUTTMAN
THE ISRAEL INSTITUTE OF APPLIED SOCIAL RESEARCH,
JERUSALEM, ISRAEL

By a simplex is meant a set of statistical variables whose interrelations reveal a simple order pattern. For the case of quantitative variables, an order model was analyzed previously which allowed only for positive correlations among the variables and a limited type of gradient among the correlation coefficients. The present paper analyzes a more general model and shows how it is more appropriate to empirical data. Among the novel features emerging from the analysis are: (a) the "factoring" implied of the correlation matrix; (b) the use of a non-Euclidean distance function; and (c) the possible underlying psychological theories.

I. *Introduction*

In a new approach to factor analysis, called *radex* theory, it has been shown (**3, 4**) how two important special cases arise: the *simplex* and the *circumplex*. Only a restricted case of the simplex was considered parametrically in (**3**), allowing only positive correlations among the observed variables and only a limited type of gradient among the correlation coefficients. The purpose of the present paper is to give a parametric theory and analysis of a more general type of simplex. In this generalization, a more flexible gradient is possible, and negative correlations can appear as well as positive ones. Thus, "inhibiting" as well as "reinforcing" factors can be considered. Generalizing the parametric system for a simplex immediately suggests analogous generalizations for the circumplex, and hence also for a complete radex. We shall consider here only the simplex, and it will be clear what the implications are for the circumplex and radex.

As in conventional factor analysis, we consider a universe of tests for a population of subjects. Both the universe and the population are usually theoretically indefinitely large, and in practice only a finite sample is drawn from each. It will be convenient to consider a finite battery of n tests from the universe, but to consider the population of testees to be infinitely large so that we need not be concerned with sampling error due to people. We shall then be able to see what happens as n increases.

A particularly curious result of the present analysis is as follows. It turns out that in terms of ordinary factor analysis, one should factor not the co-

*Read at the International Congress of Psychology, Montreal, June 7–12, 1954. This research was facilitated in part by an uncommitted grant-in-aid to the writer from the Behavioral Sciences Division of the Ford Foundation.

variance matrix of our generalized simplex, but rather the *inverse* of this matrix. The factoring implied is of two kinds. First, the *first centroid*—in the sense of Thurstone—should be factored from the *inverse* matrix, and then the *principal components* should be taken as the remaining $n - 1$ factors. This particular way of regarding the factor resolution turns out to have important theoretical and practical implications for the present simplex theory.

A second and most highly important result reveals a limitation of considering variables as points only in a Euclidean space. Regarded this way, our simplex appears n-dimensional, or with as many Euclidean dimensions as distinct variables. However, when distances between these same points are measured in a certain non-Euclidean fashion, then the points can be plotted on a straight line, or they form a *one*-dimensional non-Euclidean system.

Further novel features appear in our generalized simplex with respect to the psychological theories that can possibly account for it.

II. *General Notation*

Let t_{ij} denote the observed score of person i on test j. The mean and the standard deviation of each test are arbitrary, and indeed are usually artifacts of the test construction procedure (**3**). One part of the problem of factor analysis is to express each t_{ij} as the sum of two types of components: common and deviant (or "unique"). Let e_{ij} be the score of person i on the deviant component of test j. Then we can write, for all i and j,

$$t_{ij} = w_j s_{ij} + e_{ij} , \tag{1}$$

where s_{ij} is the structural or non-deviant part of t_{ij} , and w_j is a multiplying constant to allow for the arbitrariness of the standard deviation of the observed t_{ij} . Especially in the simplex theory to follow, the standard deviations of the s_{ij} are *not* in general arbitrary.

Since the present simplex theory is concerned only with covariances between the s_{ij} , it will be convenient to consider the mean of each to be zero,

$$\mathrm{E}_i \, s_{ij} = 0 \qquad (j = 1, 2, \cdots , n). \tag{2}$$

Various laws of deviation are possible for the e_{ij} , as pointed out in (**3**). The one assumed in conventional factor analysis is the δ-law,

$$\mathrm{cov}\,(e_j , s_k) = \mathrm{cov}\,(e_j , e_k) = 0 \qquad (j \neq k). \tag{3}$$

A well-known consequence of (3) and (1) is that

$$\mathrm{cov}\,(t_j , t_k) = w_j w_k \,\mathrm{cov}\,(s_j , s_k) \qquad (j \neq k). \tag{4}$$

According to (4), the covariance matrix of the observed tests is derived from that of the underlying s_j merely by constants of proportionality, except

for the main diagonal. Any submatrix in the one that involves no main diagonal element must have exactly the same rank as the corresponding submatrix in the other. This suggests one way of testing hypotheses about the s_{ij}, insofar as these lead to conditions on the ranks of certain submatrices.

The δ-law (3) may or may not be true in practice for a given set of data. One approach to testing it for some data is by image analysis (6). We shall be concerned here primarily with *structural* laws or theories for the s_{ij}, and the truth or falsity of the deviance law (3) is a subsequent problem to be explored ultimately with empirical data in any given case.

III. *Review of Previous Data and Theory*

Several correlation matrices published earlier in the literature by various writers have now been re-analyzed and found to form approximate simplexes. These data represent a wide variety of mental abilities and personality traits (1, 3, 4). Two examples are shown in Tables 1 and 2. One is of a battery

TABLE 1

Correlations Among Six Numerical Ability Tests*

Test	Addition	Subtraction	Multiplication	Division	Arithmetical Reasoning	Numerical Judgment
Addition	1.00	.62	.62	.54	.29	.28
Subtraction	.62	1.00	.67	.53	.38	.37
Multiplication	.62	.67	1.00	.62	.48	.52
Division	.54	.53	.62	1.00	.62	.57
Arithmetical Reasoning	.29	.38	.48	.62	1.00	.64
Numerical Judgment	.28	.37	.52	.57	.64	1.00

*From Table 2, pp. 110-112 of (11). See analysis in (3).

TABLE 2

Correlations Among Six Tests of a Certain Type of Verbal Ability*

Test	Proverbs	Vocabulary	Word Checking	Verbal Enumeration	Association	Synonyms
Proverbs	1.00	.55	.29	.24	.18	.17
Vocabulary	.55	1.00	.46	.44	.31	.24
Word Checking	.29	.46	1.00	.56	.34	.22
Verbal Enumeration	.24	.44	.56	1.00	.43	.27
Association	.18	.31	.34	.43	1.00	.45
Synonyms	.17	.24	.22	.27	.45	1.00

*Called "abstractness of verbalization" in (4, p. 13). Data from Appendix Table 1 of (12): tests 43, 45, 58, 57, 6 and 55.

of numerical ability tests, and the other is of a certain type of verbal ability tests.

From mere inspection of Tables 1 and 2, it is clear that there is some kind of order relationship within each battery of tests. In each case, the largest correlations are next to the main diagonal, and taper off to the northeast and southwest corners of the table. No other arrangement of the rows and columns of the tables, or reshuffling of the order of the variables, will yield such an apparent gradient. It is as if one could regard the variables to be points ordered along a straight line, and the correlation of one variable with another decreases as the other departs from it—in either direction—along this line.

One of the interesting new parametric properties to be developed in the simplex theory of the present paper is that simplex variables can be *literally* plotted as points along a straight line, with distances between them being strictly additive.

It has been shown in (3) that it is possible to write a factor model which will yield a gradient among correlation coefficients that has the general characteristics of the empirical ones in Tables 1 and 2 (or of the several other known empirical examples of approximate simplexes). For example, assume there are n uncorrelated factors underlying the n tests in the battery. Let x_{ic} denote the score of person i on factor x_c. It is convenient to assume also that the means of the x_{ic} are zero. Thus, the assumptions so far can be written as

$$\mathop{\mathrm{E}}_{i} x_{ic} = 0 \qquad (c = 1, 2, \cdots, n) \tag{5}$$

and

$$\mathop{\mathrm{E}}_{i} x_{ib}x_{ic} = 0 \qquad (b \neq c;\ b, c = 1, 2, \cdots, n). \tag{6}$$

Now, assume further that there is an order within the s_j and also within the x_c such that for all i and j the following factor law of formation holds:

$$s_{ij} = \sum_{c=1}^{j} x_{ic} \qquad \begin{bmatrix}\text{Additive}\\ \text{restricted}\\ \text{simplex}\end{bmatrix}. \tag{7}$$

Let σ_{s_j} and σ_{x_c} be the standard deviations of s_j and x_c, respectively, and let $\rho_{s_j s_k}$ be the coefficient of correlation between s_j and s_k. Then it has been proved from (5), (6), and (7) that

$$\sigma^2_{s_j} = \sum_{c=1}^{j} \sigma^2_{x_c} \qquad (j = 1, 2, \cdots, n) \tag{8}$$

and

$$\rho_{s_j s_k} = \sigma_{s_j}/\sigma_{s_k} \quad (j \leqq k). \tag{9}$$

According to (8), σ_{s_k} increases as k increases, so that for fixed j in (9) the right member must decrease as k departs from j. This describes a gradient in the correlation coefficients of the s_i , which when modified in the t_i by the presence of error as in (1)—say that (3) and (4) hold—can approximately give rise to observed gradients such as in Tables 1 and 2.

Another way of writing law (7) is

$$s_{ij} = s_{i,j-1} + x_{ij} . \tag{10}$$

Equality (10) asserts that s_j is the same as its predecessor s_{j-1} , except for the addition of a new factor. Interpreting Table 1 this way would imply that—apart from deviant factors of the e_i type—the subtraction test involves the same x_1 as does the addition test, but also an x_2 not called on by the addition test. The multiplication test calls on both x_1 and x_2 , but also on an x_3 , etc. A corresponding explanation would hold for the hierarchy among the verbal ability tests of Table 2.

It has been shown in (**3**) how an entirely different factor law can give rise to exactly the same type of correlation matrix as in (9). Instead of having factors x_c that are added according to (7), it is possible to write a law wherein factors are *multiplied* by each other and yet yield a hierarchy of correlations *identical* with (9). Even other laws may yield exactly the same results.

But it has also been pointed out in (**3**) that the detection and use of the simplex pattern does not at all depend on knowing whether law (7) holds or some alternative law leading to identical results. It is sufficient to determine the law of formation of the *correlation coefficients*, say such as (9), and for this the specification of an underlying law of factors such as (7) is not strictly necessary.

An important feature of a matrix with elements of the form (9) is that, if $\sigma_{s_j} \neq \sigma_{s_k}$ whenever $j \neq k$, then the matrix is nonsingular. Furthermore, the inverse of this nonsingular matrix has zero elements everywhere except in the main diagonal and in the immediately adjacent diagonals. This has profound implications for prediction problems, since the elements of the inverse matrix are the basis for the multiple regression weights for any linear multiple regression on the s_i . This also has profound implications for the internal structure of the s_i , for these vanishing elements of the inverse show that the principal components of the s_i satisfy a certain second-order linear difference equation, and hence must obey a certain general oscillatory law of formation (**2, 3**).

We now wish to generalize law (9). We shall do this in two steps. The first stage is to use a generalization of law (7) for expository purposes.

IV. *A First Parametric Generalization of the Additive Simplex*

It is clear from (9) that only positive correlations can arise from the restricted hypothesis (7). But surely there must be an order system which

would also allow for negative correlations. It is also verifiable from (9) that any tetrad, or second-order minor determinant, must vanish if all of its elements are on one side of the main diagonal of the correlation matrix (and *not* vanish if elements come from both sides of the main diagonal). Could there be an order system that does not lead to such a restrictive condition on the rank of parts of the matrix?

A generalization of (7) that does relax these restrictions somewhat is as follows. In (7), each x_c operates as an "all or none" affair, in the sense that s_j does not involve x_c whenever $c > j$. For $c > j$, then, let us assume there is an *alternative* set of factors operating, say some y_c .

Let y_{ic} be the score of person i on alternative factor y_c . For convenience, assume the means of the y_c are zero

$$\mathop{\mathrm{E}}_{i} y_{ic} = 0 \qquad (c = 1, 2, \cdots, n). \tag{11}$$

Analogous to (6), we assume the y_c to be uncorrelated with each other,

$$\mathop{\mathrm{E}}_{i} y_{ib}y_{ic} = 0 \qquad (b \neq c;\ b, c = 1, 2, \cdots, n). \tag{12}$$

We also assume y_c to be uncorrelated with x_b whenever $b \neq c$,

$$\mathop{\mathrm{E}}_{i} x_{ib}y_{ic} = 0 \qquad (b \neq c). \tag{13}$$

Let γ_c denote the covariance between x_c and y_c

$$\gamma_c = \mathop{\mathrm{E}}_{i} x_{ic}y_{ic} \qquad (c = 1, 2, \cdots, n). \tag{14}$$

No assumptions will be made here concerning the size or sign of γ_c for any c $(c = 1, 2, \cdots, n)$. Different covariances can arise from different psychological processes. For example, if x_c is an "excitatory" factor, then y_c might be an "inhibiting" factor, and the covariance γ_c might be negative. Or x_c and y_c might denote two different levels of excitation (or of inhibition) of the same type of factor, and hence γ_c might be positive.

We can now write the following generalization of law (7)

$$s_{ij} = \sum_{c=1}^{j} x_{ic} + \sum_{c=j+1}^{n} y_{ic} \qquad \begin{pmatrix}\text{First generalization}\\ \text{of additive simplex}\end{pmatrix}. \tag{15}$$

In place of (8) we now get

$$\sigma_{s_j}^2 = \sum_{c=1}^{j} \sigma_{x_c}^2 + \sum_{c=j+1}^{n} \sigma_{y_c}^2 \qquad (j = 1, 2, \cdots, n). \tag{16}$$

It is also easy to derive from (11), (12), (13), (14), and (15) that

$$\operatorname{cov}(s_j, s_k) = \sum_{c=1}^{j} \sigma_{x_c}^2 + \sum_{c=k+1}^{n} \sigma_{y_c}^2 + \sum_{c=j+1}^{k} \gamma_c \qquad (j \leqq k). \tag{17}$$

From (17) and (16)

$$\text{cov}\,(s_i\ ,\ s_k) = \sigma^2_{s_i} + \sum_{c=j+1}^{k} (\gamma_c - \sigma^2_{y_c}) \qquad (j \leqq k), \tag{18}$$

so that (9) generalizes to

$$\rho_{s_j s_k} = (\sigma_{s_j}/\sigma_{s_k}) + \left[\sum_{c=j+1}^{k} (\gamma_c - \sigma^2_{y_c})\right] \Big/ (\sigma_{s_j}\sigma_{s_k}) \qquad (j \leqq k). \tag{19}$$

Since the second terms on the right of (18) and of (19) can be negative—especially when $\gamma_c < 0$ for some or all of the c—the left members can also be negative upon occasion. Thus, law (15) allows also for possible negative correlations among the s_{ij} .

The rank condition on the correlation matrix resulting from (9) is also relaxed a bit, according to (19). To see this, it is easiest first to deal with the covariance matrix defined by (18). Taking first differences with respect to k, we see that

$$\text{cov}\,(s_j\ ,\ s_{k+1}) - \text{cov}\,(s_j\ ,\ s_k) = \gamma_{k+1} - \sigma^2_{y_{k+1}} \qquad (j \leqq k). \tag{20}$$

In the matrix of order $n \times (n-1)$ defined by the left member of (20), all submatrices with elements all on one side of the main diagonal are clearly of rank one at most, according to the right member of (20). Hence, in the $n \times n$ matrix of the elements defined by (18), all corresponding submatrices cannot be of rank greater than two. But the rank of any submatrix in $[\rho_{s_j . s_k}]$ is the same as of the corresponding submatrix in $[\text{cov}\,(s_j\ ,\ s_k)]$ since the rows and columns of one differ from those of the other only by constants of proportionality. Hence the rank of any submatrix of $[\rho_{s_j . s_k}]$ cannot be greater than two when all its elements are on one side of the main diagonal.

Formula (19) will of course allow for a closer fit to data such as in Table 1 and Table 2 than will formula (9). This may be needed especially to account for the aberration of the subtraction test from a simple gradient; apparently subtraction differs from addition and multiplication in somewhat of another manner than called for by law (7), and law (15) may be more appropriate.

V. *A Second Parametric Generalization*

A formulation like (15) is helpful in trying to understand what kinds of processes can possibly give rise to order relations among observed correlation coefficients. However, a formula like (19) can divert attention from the main consequences of having order relationships. One might be tempted to focus, for example, on the problem of estimating the γ_c and the $\sigma^2_{y_c}$ to be used in (19). Clearly, an analysis based on observed correlation coefficients alone can only hope at best to estimate the *differences* $(\gamma_c - \sigma^2_{y_c})$, and not each term separately. That is, a correlational analysis alone cannot hope to piece out all the details of a process such as (15). Even if this were possible,

there are many important things to be learned about $[\rho_{s_i,s_k}]$ that do not need specification of these details.

We shall now give the main generalization of the simplex intended in this paper. It involves no explicit use of underlying factors x_c , y_c , or any others. Its focus is on what can be learned by a correlational analysis alone.

Each of laws (7) and (15)—given also the assumptions (6), (12), and (13)—satisfies the following necessary condition

$$\mathop{\mathrm{E}}_{i}(s_{ij} - s_{ik})(s_{ik} - s_{il}) = 0 \qquad (j \leqq k \leqq l). \tag{21}$$

This is an order condition among the s_i , and yet needs no detailed specification of an underlying factor mechanism. All that is hypothesized in (21) is that the difference between an s_k and any of its *predecessors* in the sequence is uncorrelated with the difference between this same s_k and any of its *successors* in the sequence: $s_k - s_j$ is uncorrelated with $s_k - s_l$ whenever $j \leqq k \leqq l$.

An interesting immediate consequence of (21) is that we can regard the s_i not merely as points arranged in a rank order, but we can specify an *additive metric* for distances between these points. Let d_{jk} be defined by

$$d_{jk} = \mathop{\mathrm{E}}_{i}(s_{ij} - s_{ik})^2 \qquad (j, k = 1, 2, \cdots, n). \tag{22}$$

Now, we can write the identity

$$s_{ij} - s_{il} \equiv (s_{ij} - s_{ik}) + (s_{ik} - s_{il}). \tag{23}$$

Taking expectations of the squares of both sides of (23) shows that the following theorem is true.

THEOREM 1. *A necessary and sufficient condition for the order relation* (21) *to hold is that*

$$d_{jl} = d_{jk} + d_{kl} \qquad (j \leqq k \leqq l), \tag{24}$$

where d_{jk} *is defined as in* (22).

Therefore, if we define d_{jk} to be the *distance* between points s_j and s_k , this distance function is *additive* according to (24). If s_k is between s_j and s_l , then the distance from s_j to s_l is the sum of the distances from s_j to s_k and from s_k to s_l . Accordingly, the n points s_j can be plotted on a straight line, with distances between each pair being determined by formula (22).

It has been customary in factor analysis to regard all variables involved as being in a *Euclidean* space. For such a space, the distance between two points s_j and s_k is defined as the *square root* of d_{jk} . This makes the dimensionality of the space necessarily equal to the rank of the matrix $[\rho_{s_i,s_k}]$. Now this matrix is in general nonsingular when (21) holds, or n Euclidean dimensions are required. Using the non-Euclidean metric of (22) leads to but a one-dimensional space, according to Theorem 1.

It should be remarked that the distance function (22) does not yield a metric space in the general case of arbitrary variables, for the requisite triangular inequality need not be satisfied. However, we are using it here only for the special case where (21) holds, so the space of the specific points involved is certainly metric, being even one-dimensional in the sense of (24).

The writer first used a metric of the type (22) in the context of the principal components of scale analysis of qualitative data (**8**), and this suggested the developments presented here for a simplex of quantitative variables.

VI. *The Rank of Certain Submatrices*

From now on we shall be concerned largely with the covariances among the s_j , so it will be convenient to let σ_{jk} denote the covariance between s_j and s_k ,

$$\sigma_{jk} = \operatorname{cov}(s_j , s_k) = \underset{i}{\mathrm{E}}\, s_{ij}s_{ik} \qquad (j, k = 1, 2, \cdots , n). \tag{25}$$

We wish to prove the following theorem:

THEOREM 2. *If n variables s_j satisfy the order condition* (21), *then any submatrix of $[\sigma_{jk}]$ cannot be of rank greater than 2 if all its elements are on one side of, or on, the main diagonal.*

For the proof, we first expand (21), using notation (25), to obtain

$$\sigma_{jk} = \sigma_k^2 + \sigma_{jl} - \sigma_{kl} \qquad (j \leqq k \leqq l). \tag{26}$$

Since $[\sigma_{jk}]$ is a symmetric matrix, it suffices to consider only submatrices on one side of the main diagonal, say with all elements to the right of (or above) the diagonal. By differencing (26) with respect to j we see that

$$\sigma_{j+1,k} - \sigma_{jk} = \sigma_{j+1,l} - \sigma_{jl} \qquad (j + 1 \leqq k \leqq l). \tag{27}$$

According to (27), all elements to the right of the main diagonal and in the same row of the $n \times (n - 1)$ matrix $[\sigma_{j+1,k} - \sigma_{jk}]$ are equal. Hence no submatrix which is all to one side of the main diagonal can have a rank exceeding unity. Consequently, the corresponding submatrices in $[\sigma_{jk}]$ cannot have ranks greater than 2, or Theorem 2 is proved.

VII. *The Problem of Weights for Principal Components*

Related to Theorem 2, but perhaps more striking, are two laws of formation: one for the inverse matrix and one for the principal components of $[\sigma_{jk}]$ when (21) holds.

In developing these laws, we first wish to take into account the fact that the principal components of a covariance matrix depend in part on the weight functions used, or the relative sizes of the standard deviations of the variables concerned. The components may shift also as one removes

variables from the matrix, or introduces additional ones. To see the effects of these operations when (21) is true, we shall introduce further notation.

First, we shall allow for the possibility that there is a frequency distribution over our n points s_i . This can arise from the fact that any observed variables t_k are but a sample from an infinite universe of variables. While each t_k has a different e_k , many can have exactly the same s_i , or be aimed at exactly the same aspect of the underlying simplex. Let f_i be the relative frequency of s_i in this sense; that is, f_i is the proportion of all the t_k which have the same s_i in (1). Then

$$\sum_{i=1}^{n} f_i = 1. \tag{28}$$

Next, we shall allow for the possibility that it is not the σ_{jk} themselves to be analyzed, but perhaps the $\rho_{s_j s_k}$, or some other weighted function of the σ_{jk} . Let v_i be the weight associated with s_i . Thus, if the principal components of $\rho_{s_j s_k}$ are to be analyzed, then $v_i = 1/\sigma_i$. If the principal components of the $t_i - e_i$ are to be analyzed, as in (1), then $v_i = w_i$. In general, the v_i represent any set of real numbers, and we wish to know the principal components of the Gramian matrix $[v_i v_k \sigma_{jk}]$ when relative frequency f_i is associated with row and column j.

Let λ denote a latent root of the matrix, and let z_i be the jth element of the associated latent vector. Our job is to solve the stationary equations (cf. **3**)

$$\sum_{i=1}^{n} z_i f_i v_i v_k \sigma_{jk} = \lambda z_k \qquad (k = 1, 2, \cdots, n). \tag{29}$$

To simplify notation for the solution, let

$$u_i = z_i / v_i , \qquad a_i = f_i v_i^2 \qquad (j = 1, 2, \cdots, n). \tag{30}$$

Then (29) can be rewritten as

$$\sum_{i=1}^{n} a_i u_i \sigma_{jk} = \lambda u_k \qquad (k = 1, 2, \cdots, n). \tag{31}$$

It should be remarked that, from (30), the a_i are always non-negative, even though the v_i may be negative. There is no loss of generality, then, assuming all the a_i to be positive,

$$a_i > 0 \qquad (j = 1, 2, \cdots, n), \tag{32}$$

for if $a_i = 0$, this would be equivalent to $f_i = 0$, or no s_i to begin with to use in (31).

VIII. *Deriving the Inverse of the Covariance Matrix*

It will prove convenient to study (31) by means of the inverse of $[\sigma_{jk}]$. This is more than just a matter of convenience, for the inverse matrix is of

basic importance in its own right. It provides the regression coefficients in multiple correlation problems involving all the s_i , and it provides the partial correlation and multiple correlation coefficients involved. In short, it is a basic tool of image analysis (6).

Eventually one would like to know about the inverse of the observed correlation matrix $[\rho_{t_i t_k}]$. Since this will depend partly on the deviance law of the e_i in (1), all we shall do in the present paper is analyze the case where there is no error; we shall concentrate only on $[\sigma_{jk}]$. But even so, it is important to allow for the frequency function f_i , and to be concerned ultimately with the infinite universe of variables and not just a finite observed sample therefrom (6, 7).

If $[\sigma_{jk}]$ is nonsingular, let σ^{jk} denote the typical element of the inverse matrix. The inverse must be symmetric, since the covariances are. Thus,

$$\sigma^{jk} = \sigma^{kj} \qquad (j, k = 1, 2, \cdots, n). \tag{33}$$

If δ_{jk} denotes Kronecker's delta, then

$$\sum_{j=1}^{n} \sigma_{jk}\sigma^{jl} = \delta_{kl} \qquad (k, l = 1, 2, \cdots, n). \tag{34}$$

We shall solve (34) for σ^{jl} by a differencing process.

The following differencing notation will be used. If z_k , z_{jk} , or z_{kl} are any quantities to be differenced with respect to k, then

$$\underset{k}{\Delta}\, z_k = z_{k+1} - z_k , \qquad \underset{k}{\Delta}\, z_{jk} = z_{j,k+1} - z_{jk} , \qquad \underset{k}{\Delta}\, z_{kl} = z_{k+1,l} - z_{kl} . \tag{35}$$

If we let $l = k + 1$ in (26), the equations can be rewritten as

$$\underset{k}{\Delta}\, \sigma_{jk} = \begin{cases} \sigma_{k,k+1} - \sigma_k^2 & (j \leqq k) \\ \sigma_{k+1}^2 - \sigma_{k,k+1} & (j > k) \end{cases} \qquad (k = 1, 2, \cdots, n - 1). \tag{36}$$

Differencing both members of (34) with respect to k and using (36) yield

$$(\sigma_{k,k+1} - \sigma_k^2) \sum_{j=1}^{k} \sigma^{jl} + (\sigma_{k+1}^2 - \sigma_{k,k+1}) \sum_{j=k+1}^{n} \sigma^{jl} = \underset{k}{\Delta}\, \delta_{kl}$$

$$\begin{pmatrix} k = 1, 2, \cdots, n - 1 \\ l = 1, 2, \cdots, n \end{pmatrix}. \tag{37}$$

Let α_l be the sum of the elements in the lth column (row) of $[\sigma^{jl}]$,

$$\alpha_l = \sum_{j=1}^{n} \sigma^{jl} \qquad (l = 1, 2, \cdots, n). \tag{38}$$

Also, notice from (22) that

$$d_{k,k+1} = \sigma_k^2 - 2\sigma_{k,k+1} + \sigma_{k+1}^2 \qquad (k = 1, 2, \cdots, n - 1). \tag{39}$$

By bringing in the notion of a frequency function f_i we are in effect assuming our n points s_i to be distinct, or that

$$d_{k,k+1} > 0 \qquad (k = 1, 2, \cdots, n-1). \tag{40}$$

Let b_k and c_k be defined respectively as

$$b_k = (\sigma^2_{k+1} - \sigma^2_{k,k+1})/d_{k,k+1} \quad c_k = 1/d_{k,k+1} \quad (k = 1, 2, \cdots, n-1). \tag{41}$$

Then by using (38), (39) and (41), we obtain from (37) that

$$\sum_{i=1}^{k} \sigma^{il} = \alpha_l b_k - c_k \underset{k}{\Delta} \delta_{kl} \qquad \begin{pmatrix} k = 1, 2, \cdots, n-1 \\ l = 1, 2, \cdots, n \end{pmatrix}. \tag{42}$$

Taking first differences in (42) with respect to k yields the important second-order difference equation

$$\sigma^{k+1,l} = \alpha_l \underset{k}{\Delta} b_k - \underset{k}{\Delta} (c_k \underset{k}{\Delta} \delta_{kl}) \qquad \begin{pmatrix} k = 1, 2, \cdots, n-2 \\ l = 1, 2, \cdots, n \end{pmatrix}. \tag{43}$$

We now wish to obtain an explicit formula for α_l in (43). As is well known for Kronecker delta's,

$$\sum_{l=1}^{n} \delta_{kl} \equiv 1, \qquad \sum_{l=1}^{n} \underset{k}{\Delta} \delta_{kl} \equiv 0. \tag{44}$$

Hence, if we let α be the sum of all n^2 elements of σ^{ik}, or

$$\alpha = \sum_{l=1}^{n} \alpha_l = \sum_{j=1}^{n} \sum_{k=1}^{n} \sigma^{jk}, \tag{45}$$

and if we sum both members of (42) over l, we obtain

$$\sum_{i=1}^{k} \alpha_i = \alpha b_k \qquad (k = 1, 2, \cdots, n-1). \tag{46}$$

Since $[\sigma^{ik}]$ must be Gramian if it exists, the last member of (45) shows α to be a quadratic form over this nonsingular matrix, so it must be that

$$\alpha > 0. \tag{47}$$

For $k = 1$, (46) shows that

$$\alpha_1 = \alpha b_1 . \tag{48}$$

Differencing both members of (46) with respect to k shows further that

$$\alpha_k = \alpha \underset{k}{\Delta} b_{k-1} \qquad (k = 2, 3, \cdots, n-1); \tag{49}$$

and finally, for $k = n - 1$, (46) shows that $\alpha - \alpha_n = \alpha b_{n-1}$, or

$$\alpha_n = \alpha(1 - b_{n-1}). \tag{50}$$

Therefore, if we let

$$g_k = \begin{cases} b_1 & (k = 1) \\ \underset{k}{\Delta}\, b_{k-1} & (k = 2, 3, \cdots, n-1) \\ 1 - b_{n-1} & (k = n) \end{cases} \tag{51}$$

we can write our desired formula compactly as

$$\alpha_k = \alpha g_k \qquad (k = 1, 2, \cdots, n). \tag{52}$$

It also follows from (51) that

$$\sum_{k=1}^{n} g_k = 1, \tag{53}$$

so (52) cannot be used to obtain an explicit formula for α.

An explicit formula for α is easily obtained as follows. Multiply both members of (38) by σ_{kl} and sum over l. Recalling (34), (44), and (52), we see that—changing subscripts—

$$\alpha \sum_{j=1}^{n} g_j \sigma_{jk} = 1 \qquad (k = 1, 2, \cdots, n), \tag{54}$$

or

$$\alpha = 1 \Big/ \left(\sum_{j=1}^{n} g_j \sigma_{jk} \right) \qquad (k = 1, 2, \cdots, n). \tag{55}$$

Using notation (51), and shifting notation from $k + 1$ to k, we can now rewrite (43) as

$$\sigma^{kl} = \alpha g_k g_l - \underset{k}{\Delta}\, (c_{k-1} \underset{k}{\Delta}\, \delta_{k-1,l}) \qquad \begin{pmatrix} k = 2, 3, \cdots, n-1 \\ l = 1, 2, \cdots, n \end{pmatrix}. \tag{56}$$

Now, (56) gives all the elements of $[\sigma^{ik}]$ directly except for the first and last rows ($k = 1$ and $k = n$). These "boundary conditions" are obtained from (42). Setting $k = 1$ and using notation (50) show that

$$\sigma^{1l} = \alpha g_1 g_l - c_1 \underset{1}{\Delta}\, \delta_{1l} \qquad (l = 1, 2, \cdots, n). \tag{57}$$

Setting $k = n - 1$ in (42), and using (38) and (51), show that

$$\sigma^{nl} = \alpha g_n g_l + c_{n-1} \underset{n-1}{\Delta}\, \delta_{n-1,l} \qquad (l = 1, 2, \cdots, n). \tag{58}$$

IX. *The Inverse Matrix and the Ranks of Its Parts*

To see more graphically what the inverse matrix defined by (56), (57), and (58) looks like, let c_{kl} be defined, for all l, as

$$c_{kl} = \begin{cases} -c_1 \underset{1}{\Delta} \delta_{1l} & (k = 1) \\ -\underset{k}{\Delta}(c_{k-1} \underset{k}{\Delta} \delta_{k-1,l}) & (k = 2, 3, \cdots, n-1) \\ c_{n-1} \underset{n-1}{\Delta} \delta_{n-1,l} & (k = n). \end{cases} \tag{59}$$

The right member of (59) expands into the following explicit statement of the elements of $[c_{kl}]$:

$$[c_{kl}] = \begin{bmatrix} c_1 & -c_1 & & & & & \\ -c_1 & c_1 + c_2 & -c_2 & & & & \\ & -c_2 & c_2 + c_3 & & & & \\ & & & \cdots & & & \\ & & & & -c_{n-2} & c_{n-2} + c_{n-1} & -c_{n-1} \\ & & & & & -c_{n-1} & c_{n-1} \end{bmatrix}. \tag{60}$$

$[\sigma^{kl}]$ can now be regarded as the sum of two matrices, for we can write

$$\sigma^{kl} = \alpha g_k g_l + c_{kl} \qquad (k, l = 1, 2, \cdots, n). \tag{61}$$

Now, from (59) and (44)—or from (60)—

$$\sum_{l=1}^{n} c_{kl} = 0 \qquad (k = 1, 2, \cdots, n), \tag{62}$$

or the columns (rows) of $[c_{kl}]$ are linearly dependent and the matrix is singular. It has been proved in (**2**) that the latent roots of a matrix of the form of (60) must all be distinct. So although $[c_{kl}]$ is singular, it can have only one zero root and hence must be of rank $n - 1$. By inspection of (60) it is seen that all submatrices with elements all on one side of, or on, the main diagonal are either of rank zero or one.

On the other hand, the matrix $[\alpha g_k g_l]$ is at most of rank one, being the product of a vector and its transpose. Indeed, for $[\sigma_{jk}]$ to be nonsingular, $[\alpha g_k g_l]$ cannot vanish, else the right member of (61) will be left only with the singular $[c_{kl}]$. So a necessary condition for the inverse to exist is that $[\alpha g_k g_l]$ be precisely of rank one, or that $[g_k] \neq 0$. This last condition is always assured by (53).

From the conclusions of the last two paragraphs, it is apparent that Theorem 2 holds for $[\sigma^{ik}]$ as well as for $[\sigma_{jk}]$. (Indeed, the writer has an unpublished theorem that shows a general correspondence between ranks of

parts of an inverse and the corresponding parts of the original matrix for any nonsingular matrix. We have merely worked out a special case here.)

X. *Implications for Statistical Prediction*

While the one-sided rank 2 condition holds for $[\sigma^{ik}]$, it is the details that are important. In general, the elements of an inverse matrix depend on all the elements of the original matrix, and will change as the order of the matrix is increased or decreased. But in (61), the *only over-all factor that changes as variables are added to or removed from the battery is* α, or the sum of all the elements in the inverse.

A coefficient g_k , as defined by (51) and (41), depends only on the variances and covariances of s_k and its immediate neighbors s_{k-1} and s_{k+1} . A coefficient c_{kl} will vanish, according to (60), unless $l = k - 1$, k, or $k + 1$. Therefore, if an s_{n+1} is added beyond the s_n of the given simplex, none of these coefficients will change except those for s_n . Or if a point is inserted between s_k and s_{k+1} in the simplex order, this will change coefficients associated only with points in the neighborhood of this new point. Thus again, as discussed in great detail in (**3**), in the multiple linear regression of any s_k on the remaining $n - 1$ distinct variables of the simplex, the multiple regression weights and multiple correlation coefficients depend essentially on the law of neighboring of the points of the simplex.

Again, the possibility appears that s_k can be essentially as predictable from s_{k-1} and s_{k+1} as it is from all the $n - 1$ distinct variables in the simplex apart from itself. We shall now see how, under certain circumstances, σ^{kl} is determined largely by c_{kl} and hardly by $\alpha g_k g_l$.

Specifically, we shall prove the following theorem:

THEOREM 3. *If* λ_0 *is the smallest latent root of* $[\sigma_{jk}]$, *then*

$$\alpha \leqq 1 \Big/ \Big(\lambda_0 \sum_{j=1}^{n} g_j^2\Big). \tag{63}$$

If $\lambda_0 \sum_{j=1}^{n} g_j^2 \to \infty$ *as* $n \to \infty$, *then* $\alpha \to 0$.

For the proof, multiply both members of (54) by g_k , sum over k and use (53) to see that

$$\alpha \sum_{j=1}^{n} \sum_{k=1}^{n} g_j g_k \sigma_{jk} = 1. \tag{64}$$

Now, the value λ_0 is the smallest obtainable by the quadratic form on the left of (64) when the g_j are normalized, or

$$\sum_{j=1}^{n} \sum_{k=1}^{n} g_j g_k \sigma_{jk} \geqq \lambda_0 \sum_{j=1}^{n} g_j^2 . \tag{65}$$

Hence, (63) follows from (64) and (65), and the theorem is established.

In circumstances where the g_i are approximately equal among themselves, quantities of the form $g_i g_k / \sum_{i=1}^{n} g_i^2$ will be of the order of $1/n$. Then, if the smallest root λ_0 does not tend to zero with n, or if it tends to zero at a slower rate than the order of $1/n$, it follows from Theorem 3 that $[\alpha g_i g_k] \to 0$ as $n \to \infty$, or from (61), $[\sigma^{kl}] \to [c_{kl}]$.

To have $[\alpha g_k g_l] \to 0$ would be a special case of an ϵ-simplex as defined in (3). The general definition of an ϵ-simplex is essentially non-parametric for finite n, in the sense that it is concerned only with limits as $n \to \infty$. It simply states that multiple regression coefficients should tend to zero for non-neighboring tests, or elements more than one diagonal away from the main diagonal of the inverse matrix should tend to zero as n increases. The simplex defined by law (21) can, therefore, be a special kind of ϵ-simplex.

XI. *The Difference Equations for Principal Components*

Having an explicit formula such as (61) for the inverse matrix helps us also to study the principal components defined by (31). Multiply both members of (31) by σ^{kl} and sum over k to obtain—revising subscripts—

$$a_k u_k = \lambda \sum_{j=1}^{n} u_j \sigma^{jk} \qquad (k = 1, 2, \cdots, n). \tag{66}$$

Let β be defined by

$$\beta = \sum_{k=1}^{n} g_k u_k \,. \tag{67}$$

Then using (61) and (67) in (66) shows that

$$a_k u_k = \lambda \left(\alpha \beta g_k + \sum_{j=1}^{n} u_j c_{jk} \right) \qquad (k = 1, 2, \cdots, n). \tag{68}$$

Now, the summation on the right is also expressible as first- and second-order differences among the c_k . For, using (59), we see that

$$\sum_{j=1}^{n} u_j c_{jk} = \begin{cases} -c_1 \underset{1}{\Delta} u_1 & (k = 1) \\ -\underset{l}{\Delta} \left(c_{l-1} \underset{l}{\Delta} u_{l-1} \right) & (k = 2, 3, \cdots, n-1) \\ c_{n-1} \underset{n-1}{\Delta} u_{n-1} & (k = n). \end{cases} \tag{69}$$

Thus, (68) can be regarded as expressing a second-order difference equation with two first-order boundary conditions.

Strictly speaking, however, more than a second-order difference equation is involved in (68), for β depends on *all* n of the u_i , according to (67). However, if $\beta = 0$, then (68) certainly reduces to the right order. If $\beta \neq 0$, we can divide both members of (68) through by β and regard the unknown to be u_k/β instead of u_k . Since the u_k are determined only up to a constant of

proportionality in any event, this is one way of taking up this degree of freedom.

The properties of the solutions to (68) in the general case remain to be explored. The previous special case of the restricted simplex in (3) is where $g_1 = 1$ and $g_j = 0$ for $j = 2, 3, \cdots, n$. Then β in (67) is simply $\beta = g_1 u_1$ and (68) is

$$a_k u_k = \lambda \sum_{i=1}^{n} u_i (c_{ik} + \delta_{1i}\delta_{1k}\alpha g_1^2) \qquad (k = 1, 2, \cdots, n). \tag{70}$$

The matrix implied by the parentheses on the right differs from $[c_{ik}]$ in (60) merely by adding the quantity αg_1^2 to c_{11}, or the element c_1 in the first row and column. This merely changes the first boundary condition, as obtainable from (69), but leaves the rest of (69) unchanged.

The solutions to (70) have the law of oscillation discussed in (2) and (3).

Another special case of interest is where g_j is constant for all j. From (53), this implies

$$g_j = 1/n \qquad (j = 1, 2, \cdots, n). \tag{71}$$

If in addition, weights are chosen so that a_j is constant for all j, say

$$a_j = a \qquad (j = 1, 2, \cdots, n), \tag{72}$$

then it is easily seen from (68), (67) and (62) that $[g_i]$ is a latent vector with latent root $\lambda = na/\alpha$. Since all other latent vectors must be orthogonal to this one, it follows from (67) that $\beta = 0$ for the remaining latent vectors, and we are back to our standard type of difference equation for these remaining vectors; they are the vectors of $[c_{ik}]$. Hypotheses (71) and (72) lead to the case where the centroid is the same as a latent vector.

This raises the following question. If a resolution into components is desired, why not work in any case with those indicated by the formula for the inverse matrix? Certainly, basic structure properties are revealed by (61). If we again assume (72), then the first centroid loadings of (61) are $\sqrt{\alpha} g_k (k = 1, 2, \cdots, n)$. According to the general formulas of (9) and (10), *any* Gramian matrix can have its rank reduced by extracting a centroid—the process is not restricted to correlation matrices and can be used on $[\sigma^{kl}]$ in particular. If we subtract out the contributions of these loadings from $[\sigma^{kl}]$, then we are left with the matrix of rank $n - 1$, $[c_{kl}]$, which now has an interesting law of principal components.

The factoring law suggested by this, then, is first to remove the first centroid, and then resolve the rest into principal components.

Since we are not factoring $[\sigma_{ik}]$ here but its inverse, we are not factoring the observed scores. Rather, by implication we are factoring the *anti-image* scores, for $[\sigma^{kl}]$ is closely related to the covariances among the anti-images of the s_i (6). That factoring a Gramian matrix is equivalent to factoring a

score matrix of which it is the product has been proved in (**9**) and discussed also in (**10**).

If (72) does not hold, then a more general weighted average is called for than the centroid to remove the term in $\alpha g_k g_l$ from (61).

XII. *The Sufficiency of the Formula for the Inverse Matrix.*

Up until now, we have not mentioned a somewhat important question. Under what conditions does (61) provide a matrix that is actually inverse to $[\sigma_{jk}]$? We have arrived at (61) by assuming $[\sigma_{jk}]$ to be nonsingular and (40) to hold. But can $[\sigma_{jk}]$ be nonsingular if law (21) holds? Fully to establish (61), we must prove that (34) actually holds assuming that $[\sigma_{jk}]$ obeys law (21), or (36).

A first indispensable assumption clearly is that (40) holds, for if two points coincide, $[\sigma_{jk}]$ must obviously be singular. Next, let us examine the assertion in (55) that the sums of the rows of $[\sigma_{jk}]$, when weighted by the g_j, are constant. Let h_k be defined as

$$h_k = \sum_{j=1}^{n} g_j \sigma_{jk} \qquad (k = 1, 2, \cdots, n). \tag{73}$$

Differencing both members of (73) with respect to k and using (36) yield

$$\underset{k}{\Delta} h_k = (\sigma_{k,k+1} - \sigma_k^2) \sum_{j=1}^{k} g_j + (\sigma_{k+1}^2 - \sigma_{k,k+1}) \sum_{j=k+1}^{n} g_j \qquad (k = 1, 2, \cdots, n-1). \tag{74}$$

From (51) and (53),

$$\sum_{j=1}^{k} g_j = b_k, \qquad \sum_{j=k+1}^{n} g_j = 1 - b_k \qquad (k = 1, 2, \cdots, n-1). \tag{75}$$

Multiply both members of (74) by c_k, use (75) and notation (41)—remembering (39)—to see that

$$c_k \underset{k}{\Delta} h_k = (b_k - 1)b_k + b_k(1 - b_k) = 0 \qquad (k = 1, 2, \cdots, n-1). \tag{76}$$

Therefore $\Delta_k h_k = 0$ for all possible k, or the h_k are constant.

Should the constant value of h_k be zero, then $[\sigma_{jk}]$ would be singular, according to the resulting linear dependence expressed by (73). Therefore, for the inverse to exist, we must assume the constant value of h_k to be different from zero. Let us denote this constant value by $1/\alpha$, or define α by (55).

We can now go ahead to define a matrix $[\sigma^{kl}]$ by (61), and proceed to

prove that it satisfies (34). Multiply both members of (61) by σ_{jk}, sum over k, and use (55) to see that

$$\sum_{k=1}^{n} \sigma_{jk}\sigma^{kl} = g_l + \sum_{k=1}^{n} \sigma_{jk}c_{kl} \qquad (j, l = 1, 2, \cdots, n). \tag{77}$$

In (59), since $c_{kl} = c_{lk}$, interchange k and l; multiply through by σ_{jk}, and sum over k to obtain

$$\sum_{k=1}^{n} \sigma_{jk}c_{kl} = \begin{cases} -c_1 \underset{1}{\Delta} \sigma_{j1} & (l = 1) \\ -\underset{l}{\Delta} (c_{l-1} \underset{l}{\Delta} \sigma_{j,l-1}) & (l = 2, 3, \cdots, n-1) \\ c_{n-1} \underset{n-1}{\Delta} \sigma_{j,n-1} & (l = n). \end{cases} \tag{78}$$

Using (36) in (78), remembering notation (41) and (51), shows that

$$\sum_{k=1}^{n} \sigma_{jk}c_{kl} = \delta_{jl} - g_l \qquad (j, l = 1, 2, \cdots, n). \tag{79}$$

Substituting (79) into (77) shows that (34) holds, which is what was to be proved. These results can be summarized as a theorem.

THEOREM 4. *If* $[\sigma_{jk}]$ *satisfies law* (21), *then a necessary and sufficient condition for it to be nonsingular is that* $d_{k,k-1} > 0$ $(k = 1, 2, \cdots, n-1)$ *and that* $\sum_{j=1}^{n} g_j\sigma_{jk}$ *be different from zero for at least one value of* k. *Then* $[\sigma_{jk}]^{-1}$ *is given by formula* (61).

REFERENCES

1. Gabriel, R. K. The simplex structure of the Progressive Matrices test. *British J. of statist. Psychology*, 1954, **7**, 9–14.
2. Guttman, L. The principal components of scale analysis. In Samuel A. Stouffer, et al., Measurement and prediction, Princeton Univ. Press, 1950.
3. Guttman, L. A new approach to factor analysis: the radex. In Lazarsfeld, Paul F. (editor), Mathematical thinking in the social sciences, The Free Press, 1954.
4. Guttman, L. Two new approaches to factor analysis. Annual technical report on project Nonr-731(00), submitted to the Office of Naval Research, Washington, D. C., by International Public Opinion Research, Inc. and the Israel Institute of Applied Social Research, June 1953.
5. Guttman, L. The principal components of scalable attitudes. In Lazarsfeld, Paul F. (editor), Mathematical thinking in the social sciences, The Free Press, 1954.
6. Guttman, L. Image theory for the structure of quantitative variates. *Psychometrika*, 1953, **18**, 277–296.
7. Guttman, L. Some necessary conditions for common-factor analysis. *Psychometrika*, 1954, **19**, 149–161.
8. Guttman, L. An additive metric from all the principal components of a perfect scale. *British J. statist. Psychol.*, 1955, **8**, (In press).

9. Guttman, L. General theory and methods for matric factoring. *Psychometrika*, 1944, 9, 1–16.
10. Guttman, L. Multiple group methods for common-factor analysis: their basis, computation, and interpretation. *Psychometrika*, 1952, 17, 209–222.
11. Thurstone, L. L. Primary mental abilities. *Psychometric Monographs No.* 1, Univ. Chicago Press, 1938.
12. Thurstone, L. L. and Thurstone, T. G. Factorial studies of intelligence. *Psychometric Monographs No.* 2, Univ. Chicago Press, 1941.

Manuscript received 6/1/54
Revised manuscript received 9/13/54

PSYCHOMETRIKA—VOL. 20, NO. 3
SEPTEMBER, 1955

EQUATING TEST SCORES—A MAXIMUM LIKELIHOOD SOLUTION

FREDERIC M. LORD
EDUCATIONAL TESTING SERVICE

Certain problems of equating are discussed. The maximum likelihood solution is presented for the following special equating problem: Two tests, U and V, are to be equated, making use of a third "anchor" test, W. The examinees are divided into two random halves. Tests U and W are administered to one half; tests V and W are administered to the other half. It is assumed that any practice effect or other effect, exerted by U and V on W, is the same for U and for V.

Two tests may be said to be equated for a given group when the score scales on the two tests are so adjusted that both tests have the same frequency distribution of true scores in the given group. [Flanagan (**1**) and Gulliksen (**2**, pp. 296–304) give brief discussions of various methods of equating.] If the tests are equally reliable, then both tests will also have approximately the same frequency distribution of actual scores. As an approximation, two equally reliable tests may be equated by changing the score scale on either test in such a way that the distribution of actual scores becomes the same for both tests. The equipercentile method of equating is commonly used for this purpose.

If we wish to equate two equally reliable and otherwise approximately parallel forms of the same test, it is often convenient to assume that the score distributions of the two forms may differ somewhat in mean and variance, but that any other differences in the shape of these distributions may be ignored in practice. Under this assumption, the tests can be equated by simply changing the origin and the size of the unit of measurement of either score scale. If x and y are scores on two tests, the standardized scores $(x - \mu_x)/\sigma_x$ and $(y - \mu_y)/\sigma_y$ (μ and σ denote mean and standard deviation in the population of examinees for which the tests are to be equated) both have zero mean and unit variance; consequently, under the assumption outlined, standardized scores are equated, by definition.

Under the assumption of the foregoing paragraph, which will be implicit in all that follows, the only practical problem is to estimate μ_x , μ_y , σ_x , and σ_y for the population in which the two tests are to be used, so that the scores on both tests can be standardized. An obvious procedure is to administer test X to one random sample from this population and test Y to another random sample, and to estimate the desired parameters from the usual sample statistics.

The procedure just outlined, however, is not very efficient, since chance fluctuations produce differences in ability between the two groups, and these differences cause a bias in the equating. A more efficient method, provided the practice effect is properly handled, is to administer both tests to each examinee. Unfortunately, it is frequently not possible in practice to obtain sufficient testing time to administer two full-length tests to each examinee. A compromise procedure, suggested by Ledyard Tucker and commonly used at Educational Testing Service, is to divide the examinees into two random samples, each of which takes only one form of the test to be equated, and each of which also takes the same "anchor test." If this anchor test correlates highly with the other tests, its use greatly reduces the sampling errors of both the estimates obtained and the resulting equated scores. [Standard error formulas for equated scores obtained by the methods discussed here and by certain other methods are given in (**4**).]

It might be thought that the best procedure would be to equate each of the two forms to the anchor test and thus to each other. Actually, this procedure is inefficient, yielding estimates that, in certain cases, have considerably larger sampling errors than those obtained by ignoring the anchor test. An optimum equating procedure for handling the data in question is found by using the maximum likelihood method of estimation. The necessary estimates are derived, and the optimum procedure is outlined in what follows. The formulas that will be obtained differ only slightly from those used in Tucker's procedure, as discussed by Gulliksen (**2**, pp. 299–304); the assumptions made in reaching the present formulas are somewhat different.

Problem

Two tests, U and V, are to be equated, making use of a third anchor test, W. The examinees are divided into two random halves, which will be called the "a-group" and the "b-group." Tests U and W are administered to the a-group; tests V and W are administered to the b-group. It is assumed that any practice effect or other effect exerted by U and V on W is the same for U and for V.

R. S. Levine and W. Angoff (personal communication) have shown that the solution given here is also applicable when test W is a part of U and of V, i.e., tests U and V have common items W.

Notation

Consideration will be limited to the case where there are N examinees in each half-group. Let u_a and w_a denote the scores of examinee a, who is in the a-group, on tests U and W; let v_b and w_b similarly denote scores of examinee b, who is in the b-group.

The symbols μ, σ, and ρ will be used to represent means, standard deviations, and correlation coefficients, respectively, in the population. The

population referred to here and in what follows is the population of all examinees from which the a- and b-group may be considered to be random samples.

Sample means will be denoted by $\bar{u}$, $\bar{v}$, $\bar{w}$; sample standard deviations and correlations by s and r with appropriate subscripts. Where the meaning would otherwise be unclear, a single prime or a double prime will denote a statistic from the a-group or from the b-group, respectively.

Assumption

It is assumed that the scores u, v, and w have a normal trivariate distribution in the population. The joint distribution of u_a and w_a is thus

$$f_a(u_a , w_a) = \frac{1}{2\pi\sigma_u\sigma_w\sqrt{1-\rho_u^2}} \exp\left[-\frac{1}{2(1-\rho_u^2)}\left\{\frac{(u_a-\mu_u)^2}{\sigma_u^2} + \frac{(w_a-\mu_w)^2}{\sigma_w^2} - 2\rho_u\frac{(u_a-\mu_u)}{\sigma_u}\frac{(w_a-\mu_w)}{\sigma_w}\right\}\right], \qquad (1)$$

ρ_u being the correlation between u and w. The joint distribution of v_b and w_b, denoted by $f_b(v_b , w_b)$, is the same as the foregoing except that u is replaced by v and a by b.

The Likelihood Function

The *likelihood* of occurrence of the actually observed values of u_a and w_a in the a-group is, by definition, $\prod_{a=1}^{N} f_a(u_a , w_a)$. Similarly, the *likelihood* for the b-group is $\prod_{b=1}^{N} f_b(v_b , w_b)$. The product of these two is the *likelihood function* (L) for all observed values in the data at hand. It will be convenient to work with the logarithm of the likelihood function, which is readily found to be

$$\begin{aligned}\log L = {} & -2N\log 2\pi - N\log\sigma_u\sigma_v \\ & - 2N\log\sigma_w - \tfrac{1}{2}N\log(1-\rho_u^2)(1-\rho_v^2) \\ & - \frac{1}{2(1-\rho_u^2)}\left[\frac{1}{\sigma_u^2}\sum_a(u_a-\mu_u)^2 + \frac{1}{\sigma_w^2}\sum_a(w_a-\mu_w)^2 \right. \\ & \left. - \frac{2\rho_u}{\sigma_u\sigma_w}\sum_a(u_a-\mu_u)(w_a-\mu_w)\right] - \frac{1}{2(1-\rho_v^2)}\left[\frac{1}{\sigma_v^2}\sum_b(v_b-\mu_v)^2 \right. \\ & \left. + \frac{1}{\sigma_w^2}\sum_b(w_b-\mu_w)^2 - \frac{2\rho_v}{\sigma_v\sigma_w}\sum_b(v_b-\mu_v)(w_b-\mu_w)\right]. \qquad (2)\end{aligned}$$

The likelihood function contains eight unknown population parameters: $\mu_u , \mu_v , \mu_w , \sigma_u , \sigma_v , \sigma_w , \rho_u , \rho_v$. We wish to choose values of these parameters that will maximize the likelihood of occurrence of the actually observed sample. Consequently, we differentiate (2) with respect to each parameter

in turn and set each derivative equal to zero, at the same time placing a circumflex above the symbol for each parameter to indicate that we are now dealing with estimates of the parameters rather than with their true values. Eight simultaneous equations in eight unknowns are thus obtained.

The Likelihood Equations

After some cancellation and rearrangement, the first three equations are

$$\hat{\mu}_u - \hat{\beta}_{uw}\hat{\mu}_w = \bar{u} - \hat{\beta}_{uw}\bar{w}', \tag{3}$$

$$\hat{\mu}_v - \hat{\beta}_{vw}\hat{\mu}_w = \bar{v} - \hat{\beta}_{vw}\bar{w}'', \tag{4}$$

$$\frac{\hat{\mu}_w - \hat{\beta}_{wu}\hat{\mu}_u}{\hat{\kappa}_u^2} + \frac{\hat{\mu}_w - \hat{\beta}_{wv}\hat{\mu}_v}{\hat{\kappa}_v^2} = \frac{\bar{w}' - \hat{\beta}_{wu}\bar{u}}{\hat{\kappa}_u^2} + \frac{\bar{w}'' - \hat{\beta}_{wv}\bar{v}}{\hat{\kappa}_v^2}, \tag{5}$$

where $\hat{\kappa}_u^2 = 1 - \hat{\rho}_u^2$, $\hat{\kappa}_v^2 = 1 - \hat{\rho}_v^2$, and each $\hat{\beta}$ is a regression coefficient—for example, $\hat{\beta}_{uw} = \hat{\sigma}_u\hat{\rho}_u/\hat{\sigma}_w$.

Multiplying (3) by $\hat{\beta}_{wu}/\hat{\kappa}_u^2$, multiplying (4) by $\hat{\beta}_{wv}/\hat{\kappa}_v^2$, and adding both products to (5), we obtain, after simplification,

$$\hat{\mu}_w = \bar{w}, \tag{6}$$

where $\bar{w} = \frac{1}{2}(\bar{w}' + \bar{w}'')$ is the observed mean of w in the combined a- and b-group. Equation 6 presents the maximum likelihood estimate of μ_w.

Substituting (6) in (3) and in (4), we obtain, after simplification,

$$\hat{\mu}_u = \bar{u} - \hat{\beta}_u D, \tag{7}$$

$$\hat{\mu}_v = \bar{v} + \hat{\beta}_v D, \tag{8}$$

where $D = \frac{1}{2}(\bar{w}' - \bar{w}'')$, $\hat{\beta}_u$ is written for $\hat{\beta}_{uw}$, and $\hat{\beta}_v$ for $\hat{\beta}_{vw}$. These equations will be of practical use as soon as expressions have been found for $\hat{\beta}_u$ and $\hat{\beta}_v$.

The remaining five maximum likelihood equations are readily found to be

$$\hat{\sigma}_u^2 = (S_u^2 - \hat{\beta}_u C_{uw})/\hat{\kappa}_u^2, \tag{9}$$

a similar equation for v instead of u,

$$2\hat{\sigma}_w^2 - \frac{1}{\hat{\kappa}_u^2}\left(S_w'^2 - \frac{1 - \hat{\kappa}_u^2}{\hat{\beta}_u} C_{uw}\right) - \frac{1}{\hat{\kappa}_v^2}\left(S_w''^2 - \frac{1 - \hat{\kappa}_v^2}{\hat{\beta}_v} C_{vw}\right) = 0, \tag{10}$$

$$\hat{\rho}_u - \frac{\hat{\rho}_u}{\hat{\kappa}_u^2}\left[\frac{S_w'^2}{\hat{\sigma}_w^2} + \frac{1}{\hat{\sigma}_u^2}(S_u^2 - 2\hat{\beta}_u C_{uw})\right] + \frac{C_{uw}}{\hat{\sigma}_u\hat{\sigma}_w} = 0, \tag{11}$$

and a fifth equation like (11) but with v instead of u. In the foregoing equations,

$$S_u^2 = \sum_a (u_a - \hat{\mu}_u)^2/N, \tag{12}$$

$$C_{vw} = \sum_b (v_b - \hat{\mu}_v)(w_b - \hat{\mu}_w)/N, \tag{13}$$

and so forth.

Multiply (9) by $\hat{\rho}_u/\hat{\sigma}_u^2$ and subtract from (11) to obtain the result

$$\frac{\hat{\rho}_u}{\hat{\kappa}_u^2}\left(\frac{S_w'^2}{\hat{\sigma}_w^2} - \frac{\hat{\beta}_u C_{uw}}{\hat{\sigma}_u^2}\right) - \frac{C_{uw}}{\hat{\sigma}_u\hat{\sigma}_w} = 0. \tag{14}$$

Multiply (14) by $\hat{\kappa}_u^2$, write out $\hat{\kappa}_u^2$ and $\hat{\beta}_u$ in terms of $\hat{\rho}_u$, and simplify to obtain

$$S_w'^2\hat{\sigma}_u\hat{\rho}_u = C_{uw}\hat{\sigma}_w . \tag{15}$$

This may be rewritten

$$\hat{\beta}_u = C_{uw}/S_w'^2 . \tag{16}$$

By a well-known formula, (12) can be rewritten

$$S_u^2 = s_u^2 + (\bar{u} - \hat{\mu}_u)^2, \tag{17}$$

where $s_u^2 = \sum_a (u_a - \bar{u})^2/N$ is the observed standard deviation of u. From (17) and (7),

$$S_u^2 = s_u^2 + \hat{\beta}_u^2 D^2. \tag{18}$$

Similarly,

$$S_w'^2 = s_w'^2 + D^2, \tag{19}$$

$$C_{uw} = c_{uw} + \hat{\beta}_u D^2, \tag{20}$$

and so forth, where $c_{uw} = \sum_a (u_a - \bar{u})(w_a - \bar{w}')/N$ is the observed covariance of u and w in the a-group.

Substituting (19) and (20) into (16), we find after simplification

$$\hat{\beta}_u = c_{uw}/s_w'^2 . \tag{21}$$

The expression on the right is the observed regression coefficient of u on w in the a-group, so we may write finally

$$\hat{\beta}_u = b_{uw} . \tag{22}$$

From (9), (18), (20), and (22)

$$\hat{\sigma}_{u\cdot w}^2 = (s_u^2 - b_{uw}c_{uw}) = s_u^2(1 - r_{uw}^2) = s_{u\cdot w}^2 , \tag{23}$$

where $\hat{\sigma}_{u\cdot w}^2 = \hat{\sigma}_u^2\hat{\kappa}_u^2$, and $s_{u\cdot w}^2$ is the observed standard error of estimate in the a-group.

Finally, substitute (16) into (10) and simplify to obtain

$$\hat{\sigma}_w^2 = \tfrac{1}{2}(S_w'^2 + S_w''^2) = s_w^2 , \tag{24}$$

where s_w^2 is the observed variance of w in the combined a- and b-group, i.e.,

$$s_w^2 = [(\sum_a w_a^2 + \sum_b w_b^2)/2N] - \bar{w}^2. \tag{25}$$

The writer is indebted to William H. Angoff for this simplified proof of (24).

The Maximum Likelihood Estimates

The set of eight equations, (3), (6), (22), (23), (24), and three equations in v analogous to those in u, is sufficient for the practical calculation of the maximum likelihood estimates of all the unknown parameters. A more convenient set of eight equations, readily derived from these, is

$$\hat{\mu}_w = \bar{w}, \tag{26}$$

$$\hat{\mu}_u = \bar{u}' + b'_{uw}(\bar{w} - \bar{w}'), \tag{27}$$

$$\hat{\mu}_v = \bar{v}'' + b''_{vw}(\bar{w} - \bar{w}''), \tag{28}$$

$$\hat{\sigma}^2_w = s^2_w , \tag{29}$$

$$\hat{\sigma}^2_u = \hat{\sigma}^2_{u \cdot w} + \hat{\beta}^2_u \hat{\sigma}^2_w = s'^2_u + b'^2_{uw}(s^2_w - s'^2_w), \tag{30}$$

$$\hat{\sigma}^2_v = s''^2_v + b''^2_{vw}(s^2_w - s''^2_w), \tag{31}$$

$$\hat{\sigma}_{uw} = \hat{\sigma}_u \hat{\sigma}_w \hat{\rho}_u = \hat{\beta}_u \hat{\sigma}^2_w = b'_{uw} s^2_w , \tag{32}$$

$$\hat{\sigma}_{vw} = b''_{vw} s^2_w , \tag{33}$$

where $\hat{\sigma}_{uw}$ and $\hat{\sigma}_{vw}$ are estimates of the population covariances. In the foregoing eight equations, primes or double-primes have been attached for the sake of clarity to all sample values except $\bar{w}$ and s_w , these last two values being calculated from the combined a- and b-group.

(The maximum likelihood estimators presented in equations 26–33 constitute the solution of a general problem in estimating population parameters from incomplete data. A discussion of these results from this general point of view has been submitted for publication elsewhere.)

Equating

Granting the assumptions made from the start, a good equation for equating tests U and V is

$$(v - \mu_v)/\sigma_v = (u - \mu_u)/\sigma_u , \tag{34}$$

or, after rearranging,

$$v = Au + B, \tag{35}$$

where

$$A = \sigma_v/\sigma_u , \tag{36}$$

$$B = \mu_v - A\mu_u . \tag{37}$$

In (36) and (37), A and B are expressed in terms of the population parameters, which are unknown. We wish to use maximum likelihood estimates of A and B in (35). Since the maximum likelihood estimate of a certain

function of the parameters is the same as that function of the maximum likelihood estimates of the parameters, the equation to use for equating is

$$v = \hat{A}u + \hat{B}, \tag{38}$$

where $\hat{A} = \hat{\sigma}_v/\hat{\sigma}_u$, $\hat{B} = \hat{\mu}_v - \hat{A}\hat{\mu}_u$, the values of $\hat{\mu}_u$, $\hat{\mu}_v$, $\hat{\sigma}_u$, and $\hat{\sigma}_v$ being computed from the data by means of equations 27, 28, 30, and 31.

The formulas for equating thus obtained by the maximum likelihood method differ from those of Tucker, as discussed by Gulliksen, chiefly as a result of the fact that Tucker's procedure calls for estimating the performance of the b-group on test U, whereas the present procedure calls for estimating the performance of the entire population on both tests U and V. The present development is based on the assumption that the two groups tested are random samples from the same population. The assumptions made in Tucker's development do not require this, but they do impose considerable restriction on the nature of the differences between the two groups.

Numerical Example

The following illustrative example is based on real data taken from Karon's empirical study of equating methods (3). The raw data are given in the top half of Table 1; the necessary maximum likelihood estimates,

TABLE 1

Raw Data and Maximum Likelihood Estimates Needed for Equating

	Group a		Group b		Combined groups
	Test U	Test W	Test V	Test W	Test W
Mean ($\bar{u}$, $\bar{v}$, $\bar{w}$)	117.85	34.36	115.33	33.42	33.89
Variance (s^2)	1129.62	116.81	1109.65	114.89	116.07
Regression on w	2.6744		2.6479		
$\hat{\mu}$	116.59		116.58		
$\hat{\sigma}^2$	1124.34		1117.92		

computed by equations (27), (28), (30), and (31), are given in the bottom half. Each group contains a random sample of 600 examinees. The final equation, obtained from (38),

$$v = .997u + 0.32, \tag{39}$$

gives the raw score (v) on test V that is equivalent to any given score (u) on test U.

If test W had not been administered, the final equation would have been

$$v = (s_v/s_u)u + \bar{v} - (s_v/s_u)\bar{u}$$
$$= .991u - 1.47. \tag{40}$$

The use of test W provides the information that the b-group is probably slightly less competent and slightly less variable than the a-group (these differences having arisen solely because of sampling fluctuations). The maximum likelihood estimates in Table 1 and the resulting equation 39 take this sampling fluctuation into account, whereas equation 40 does not.

REFERENCES

1. Flanagan, J. C. Units, scores and norms. In E. F. Lindquist (Ed)., Educational measurement. Washington, D. C.: American Council on Education, 1950, pp. 695–763.
2. Gulliksen, H. Theory of mental tests. New York: Wiley, 1950.
3. Karon, B. P. The stability of equated test scores. Research Bulletin 54-25. Princeton, N. J.: Educational Testing Service, 1954. (Multilithed)
4. Lord, F. M. Notes on comparable scales for test scores. Research Bulletin 50-48. Princeton, N. J.: Educational Testing Service, 1950. (Multilithed)

Manuscript received 5/12/54
Revised manuscript received 8/26/54

PSYCHOMETRIKA—VOL. 20, NO. 3
SEPTEMBER, 1955

AXIOMS OF A THEORY OF DISCRIMINATION LEARNING*

FRANK RESTLE
STANFORD UNIVERSITY

Analysis of an empirical theory into a formal system with specified primitive notions and axioms has the advantage of making it clear what deductions from the theory are permissible, and clarifying the internal structure of the theory. An example of such analysis is presented in this paper.

Learning theories recently published by Estes and his associates (**3**, **4**, **5**) and Bush and Mosteller (**1**) have been characterized by mathematical formulation and reasoning. The writer has offered a similar theory designed for the analysis of two-choice discrimination learning. This new theory, using a strong simplifying assumption, yields several empirical predictions which have, in the main, been verified (**9**).

According to this theory the subject is faced on each trial with a collection of cues; some are relevant to getting reward and others irrelevant. On each trial of training some relevant cues are newly conditioned to the correct response and some irrelevant cues are newly adapted. A conditioned cue contributes to a correct response. An adapted cue becomes non-functional and does not directly affect the choice reaction.

The probability that a relevant cue will be conditioned on any trial (given that it has not been conditioned on a previous trial) will be denoted by θ. Since θ is constant from trial to trial and the same for all cues, the learning functions here are the same as the conditioning functions in the work of Estes and his associates (**3**, **5**, and the "equal-θ approximation" case in **4**.)

The fundamental assumption of the theory deals with θ. This assumption is that θ is the relative weight of relevant cues in the problem. The more relevant cues there are in the problem, the greater is the probability that any given relevant cue will be conditioned and that any given irrelevant cue will be adapted. By this simplifying assumption it is possible to make the theory unusually determinate.

In the earlier paper on this theory (**9**) a number of quantitative empirical

*This paper is adapted from part of a Ph.D. dissertation submitted to the Department of Psychology, Stanford University, in November 1953. The author wishes to express his appreciation to Dr. Patrick Suppes, who guided the analysis reported in this paper. The author is now with the Human Resources Research Office, The George Washington University.

laws were developed which were tested against experimental data. In general, the proposed laws were verified.

In the present paper a more precise and complete statement of the theory is made. Using only terms definable within the language of set-theory and logic, a complete list of primitive notions is given and the axioms are stated. Deductions are carried out entirely by the methods of formal mathematics, without recourse to psychological intuition or "good sense."

Before presenting the system it may be useful to describe the mathematical notions to be used. A *binary relation* is a relation between two entities. By a *set* is meant any arbitrary collection of things. In the formula $f(x, y) = z$, the term z is the value of the binary function f. If z is a real number, we say that f is *real-valued*. An *ordered couple* is a set which has two members, with the restriction that specifying the set requires not only naming the members but also indicating what order they come in. If $\langle x, y \rangle$ is an ordered couple and $x \neq y$, then $\langle x, y \rangle \neq \langle y, x \rangle$.

The usual set-theory notation is used; if X and Y are sets, $X \cup Y$ includes everything which is in either X or Y, $X \cap Y$ includes the elements which are in both X and Y, and $X - Y$ includes all the elements which are in X and are not in Y. The empty or null set is called Λ. In the body of the paper, capital letters are used to denote the sets and the one relation used; lower-case letters designate functions, integers, and variables. One function is given the designation θ to follow earlier usage (**4**, **5**).

Primitive Notions

This system of discrimination learning is based on seven primitive notions, K, S^*, Q, w, c, a, and p. K is a set, S^* is a set of ordered couples, Q is a binary relation, w is a unary real-valued function, and c, a, and p are binary real-valued functions.

The set K is intended to be interpreted as the collection of cues. A cue is anything, concrete or abstract, present, past or future, of any description, to which the subject can learn to make a differential response. Obviously, at any given time there are cues to which the subject does not make responses —otherwise, there would be no learning. But if the subject can learn a differential response to something, by some training method, then that thing is a cue. Some cues are relatively simple energy sources. Some subjects can learn to respond to spatial or temporal patterns of objects or events; some produce reactions, overt, perceptual, or "thinking," which they can discriminate. Accounts of mediating processes can be found in work by Lawrence (**6**) and Wyckoff (**10**).

The set S^* is intended to be interpreted as any collection of two-choice discrimination problems, all of which involve the same pair of choice reactions. A problem S is uniquely associated with a pair of sets of cues: the set of

relevant cues, R, which the subject can use to predict reward, and the set of irrelevant cues, I, which are uncorrelated with reward and therefore cannot be used to predict reward.

If S is a problem in S^* and n is a positive integer, then SQn is interpreted as the statement, "problem S appears on trial n." This is true if the subject must make a choice reaction in problem S on the nth trial.

If k is a cue, $w(k)$ is interpreted as the weight of cue k. According to Axiom $D2$, w is a discrete probability distribution defined over the class K of cues.

If k is a cue and n is a positive integer, then $c(k, n)$ is the probability that k is conditioned to the correct response at the beginning of the nth trial. If k is a cue then $a(k, n)$ is the probability that k is adapted at the beginning of the nth trial.

Before stating the axioms of this system we define $\theta(S)$ as the relative weight of relevant cues in problem S. This term will appear later in the learning functions of Axiom $D7$.

Definition: If $S = \langle R, I \rangle$ is in S^*, then $\theta(S) = \sum_{k \in R} w(k) / \sum_{k \in (R \cup I)} w(k)$.

Axioms

Definition: A system $\langle K, S^*, Q, w, c, a, p \rangle$ satisfying *Axioms* $D1$—$D8$ is called a *system of simple discrimination learning.*

Axiom $D1$. K and S^* are non-empty, at most denumerable sets.

Axiom $D2$. If k is in K, $w(k) \geq 0$, and $\sum_{k \in K} w(k) = 1$.

Axiom $D3$. If $S = \langle R, I \rangle$ is in S^*, then R and I are subsets of K.

Axiom $D4$. If $S = \langle R, I \rangle$ is in S^*, then the intersection of R and I is empty.

Axiom $D5$. If S_1 and S_2 are distinct members of S^*, if n is a positive integer, and if S_1Qn, then *not* S_2Qn.

Axiom $D6$. If $S = \langle R, I \rangle$ is in S^*, then for all k in $R \cup I$, $c(k, 1) = a(k, 1) = 0$.

Axiom $D7$. If $S = \langle R, I \rangle$ is in S^* and n is a positive integer and SQn, then:

If k is in R, then $c(k, n + 1) = c(k, n) + \theta(S)[1 - c(k, n)]$ and $a(k, n + 1) = a(k, n)$.
If k is in I, $c(k, n + 1) = c(k, n)$ and $a(k, n + 1) = a(k, n) + \theta(S)[1 - a(k, n)]$.
Otherwise, $c(k, n + 1) = c(k, n)$ and $a(k, n + 1) = a(k, n)$.

Axiom $D8$. If $S = \langle R, I \rangle$ is in S^* and n is a positive integer, then

$$p(S, n) = \frac{1}{2} \cdot \frac{\sum_{k \in (R \cup I)} w(k) - \sum_{k \in I} a(k, n) \cdot w(k) + \sum_{k \in R} c(k, n) \cdot w(k)}{\sum_{k \in (R \cup I)} w(k) - \sum_{k \in I} a(k, n) \cdot w(k)}.$$

Axiom $D1$ eliminates the trivial case in which either there are no cues or there is no problem and avoids mathematical difficulties by keeping K and S^* denumerable at most. Axiom $D2$ states that w is a discrete probability function. Axiom $D3$ states that the relevant and irrelevant cues in any problem are cues in the class K. Axiom $D4$ states that no cue can be both relevant and irrelevant in the same problem. Axiom $D5$ states that only one problem may occur on a given trial. Axiom $D6$ states that the system deals with a theoretically "naive" subject who, at the beginning of training

(trial 1), had neither conditioned nor adapted to any of the cues involved. Axiom $D7$ states the laws of conditioning and adaptation, which are discussed above and in the earlier paper on this subject (**9**). Axiom $D8$ states the "law of performance," giving p, the probability of a correct response, as a function of the number of conditioned and adapted cues. Inspection will show that p is the proportion of non-adapted (i.e., still-functional) cues which are conditioned plus one-half the proportion of non-adapted cues which are unconditioned.

Theorems

The theorems to be proved could not be proved rigorously with the system in (**9**). The equations derived in Theorems 2, 3, and 4 were compared directly with experiments.

The first empirical problem of the theory is the evaluation of the learning constant, $\theta(S)$, from discrimination learning data. This is accomplished by Theorem 1, which gives an explicit function relating $p(S, n)$ to $\theta(S)$. It is found in Corollary 1.1 that $p(S, n)$ is monotonic with respect to both $\theta(S)$ and n. Therefore, graphs can be constructed to determine θ knowing the empirical learning function, which corresponds to $p(S, n)$. Since such curve-fitting is unsatisfactory when dealing with individual subjects, and is invalid when dealing with groups of subjects who have different learning constants, we derive an explicit function relating the total number of errors expected, $\sum_{n=1}^{\infty} [1 - p(S, n)]$, to $\theta(S)$. Thus, if the learning experiment is continued until the subject has achieved a high criterion, the total errors made can be used to determine $\theta(S)$. Theorem 1 and its corollaries make it possible to evaluate $\theta(S)$ in practice.

The second empirical problem has to do with the combination of cues. Experimentally, we observe representative subjects learning to discriminate between, say, black and white, and from this we determine $\theta(S_{B-W})$. In the same apparatus we observe a second group of subjects learning to discriminate, for example, high and low pitches, and we determine $\theta(S_{H-L})$. The two sets of cues, brightness and pitch, are selected as ones which will not probably affect one another perceptually. In the same apparatus a third problem is run in which both brightness and pitch cues are relevant; for example, the subject must discriminate black and high pitch from white and low pitch. Theorem 2 makes it possible to predict $\theta(S_{B \text{ and } H-W \text{ and } L})$, and thus predict performance on this combined-cues problem (**2**).

The third empirical problem has to do with transfer of training from an easy discrimination problem to a more difficult one of the same sort. For example, a subject may be trained to approach *black* and avoid *white*, and is then trained to approach *dark gray* and avoid *light gray*. The experiment is interpreted as follows: we assume that the two problems present the same cues; the difference is that some of the cues which are relevant in the easier

problem are irrelevant in the more difficult problem. The more difficult problem is constructed from the easier one by shifting some cues from the set of relevant cues into the set of irrelevant cues. To predict transfer performance we first determine the θ values of the two problems by running them separately with naive subjects. From these values and knowledge of the number of trials of training on the easy problem, we can predict $p(S_{\text{hard}}, n)$ for all trials on the hard problem (**7, 9**). The required formula is derived in Theorem 3, and the total number of errors made in transfer is derived in Corollary 3.1.

The "converse" of the experiment discussed in Theorem 3 is an experiment in which the subject is first trained on the difficult problem and is then transferred to an easier one of the same sort (**9**). The formula for predicting performance, based on knowledge of θ(easy), θ(hard), and the number of pretraining trials on the hard problem, is given in Theorem 4.

Theorems 2, 3, and 4 make exact quantitative predictions of expected performance curves. Testing the predictions against empirical results does not involve curve-fitting and the use of arbitrary empirical constants. The predicted curve can in principle be drawn before any subjects are run on the test problem, and the theory is not confirmed unless the test performance corresponds to a particular learning curve predicted.

Since the proofs of the theorems are elementary in principle and somewhat tedious, only the method of proof will be given. The careful reader can verify for himself that entirely formal proofs are possible.

THEOREM 1. *If S is in S^* and SQj for all positive integers $j \leq n$, then [using θ as an abbreviation for $\theta(S)$]*

$$p(S, n) = 1 - \tfrac{1}{2}[(1 - \theta)^{n-1}]/[\theta + (1 - \theta)^n].$$

PROOF. We note that if k is in R, $c(k, n) = 1 - (1 - \theta)^{n-1}$ and if k is in I, $a(k, n) = 1 - (1 - \theta)^{n-1}$. The theorem is obtained by elementary algebra: the above values are substituted into Axiom $D8$, all terms are divided by $\sum_{k \in (R \cup I)} w(k)$, and the definition of θ is employed to simplify.

COROLLARY 1.1. *Under the conditions of Theorem* 1, *$p(S, n)$ is a monotonic non-decreasing function of n and a monotonic increasing function of θ.*

PROOF. This follows immediately from the theorem.

COROLLARY 1.2. *Under the above conditions,*

$$\sum_{n=1}^{\infty} [1 - p(S, n)] \leqq \tfrac{1}{2} + \tfrac{1}{2}[\log \theta]/[(1 - \theta) \log (1 - \theta)].$$

PROOF. We first estimate $p(S, n)$ by the continuous function $p'(S, t) = 1 - \tfrac{1}{2}[(1 - \theta)^{t-1}]/[\theta + (1 - \theta)^t]$, and integrate $1 - p'(S, t)$ by using the substitution, $y = (1 - \theta)^t$.

THEOREM 2. *If* $S_1 = \langle R_1 , I_1 \rangle$, $S_2 = \langle R_2 , I_2 \rangle$ *and* $S_3 = \langle R_3 , I_3 \rangle$ *are in* S^* *and if* $\sum_{k \in I_1} w(k) = \sum_{k \in I_2} w(k) = \sum_{k \in I_3} w(k)$, *and if* $\sum_{k \in R_1} w(k) + \sum_{k \in R_2} w(k) = \sum_{k \in R_3} w(k)$, *then*

$$1 - \theta(S_3) = [1 - \theta(S_1)][1 - \theta(S_2)]/[1 - \theta(S_1)\theta(S_2)].$$

PROOF. The proof follows immediately from the definition of θ.

THEOREM 3. (i) *If* $S_1 = \langle R_1 , I_1 \rangle$ *and* $S_2 = \langle R_2 , I_2 \rangle$ *are in* S^* *and if* R_2 *is a subset of* R_1 *and* I_1 *is a subset of* I_2 , *and if*

$$\sum_{k \in (R_1 \cup I_1)} w(k) = \sum_{k \in (R_2 \cup I_2)} w(k), \qquad \textit{and}$$

(ii) *if for all* $i \leq n$, $S_1 Q i$, *and for all* $n + j$, $S_2 Q(n + j)$, *then*

$$p(S_2 , n + j) = \frac{\theta_2 + \frac{1}{2}(1 - \theta_2)^{i-1}[\theta_1 - \theta_2 + (1 - \theta_1)^{n+1} - \theta_2(1 - \theta_1)^n]}{\theta_2 + (1 - \theta_2)^{i-1}[\theta_1 - \theta_2 + (1 - \theta_1)^{n+1}]}.$$

PROOF. Let k be a cue in R_2 . Since R_2 is a subset of R_1 , k is also in R_1 . At the beginning of trial $n + 1$, for all k in R_1 , $c(k, n + 1) = 1 - (1 - \theta_1)^n$. After $j - 1$ further trials on the second problem, $c(k, n + j) = 1 - (1 - \theta_1)^n (1 - \theta_2)^{j-1}$. This is the conditioning of all cues relevant in the second problem. At the beginning of trial $n + 1$, for all cues in I_1 , $a(k, n + 1) = 1 - (1 - \theta)^n$. For cues which are in I_2 but are not in I_1 , $a(k, n + 1) = 1 - (1 - \theta)^0 = 0$. (The fact that these latter cues have been conditioned is of no importance, since they are not relevant.) The theorem is obtained by using Axiom *D*7 to determine $c(k, n + j)$ and $a(k, n + j)$, substituting these values into Axiom *D*8, dividing by $\sum_{k \in (R \cup I)} w(k)$, and collecting like terms.

COROLLARY 3.1. *Under the conditions of Theorem* 3,

$$\sum_{j=1}^{\infty} [1 - p(S_2 , n + j)] \doteqdot \frac{B - A}{\theta_2 + B} + \frac{B - A}{B \log (1 - \theta_2)} [\log \theta_2 - \log (\theta_2 + B)],$$

where $A = \frac{1}{2}[\theta_1 - \theta_2 + (1 - \theta_1)^{n+1} - \theta_2(1 - \theta_1)^n]$, *and* $B = \theta_1 - \theta_2 + (1 - \theta_1)^{n+1}$.

PROOF. Note that by Theorem 3,

$$1 - p(S, n + j) = [\theta_2 + (B - A)(1 - \theta_2)^{j-1}]/[\theta_2 + B(1 - \theta_2)^{j-1}].$$

This is approximated by a continuous function, substituting the real variable t for j, and the resulting function is integrated, giving the corollary.

THEOREM 4. *Given the same conditions as under* (i) *in Theorem* 3, *but if for all* $i \leq n$, $S_2 Q i$ *and if for all* $n + j$, $S_1 Q(n + j)$, *then*

$$p(S_1 , n + j) = \frac{\theta_1 - \frac{1}{2}(1 - \theta_1)^{i-1}[\theta_1 - \theta_2 + \theta_2(1 - \theta_2)^n - (1 - \theta_2)^n(1 - \theta_1)]}{\theta_1 + (1 - \theta_2)^n(1 - \theta_1)^i}.$$

PROOF. The proof is similar to that of Theorem 3.

Discussion

Certain characteristics of the axiom system offered in this paper may require explanation. The extremely abstract nature of the axioms is designed to separate carefully the formal system from its psychological interpretation. This separation makes it possible to be sure that all needed assumptions have been explicitly stated. Axioms $D1$—$D6$ are formal in nature and do not represent crucial psychological assumptions. However, if the required theorems are to be proved rigorously, such axioms are necessary (**8**).

The purpose of the paper is to make clear the formal assumptions, not their empirical consequences. However, it may be noted that if the four primitive notions K, S^*, Q, and w are defined operationally, the other three, c, a, and p, can be defined explicitly by using Axioms $D7$ and $D8$ as definitions. If the notion of a cue can be made clear, there is not likely to be any difficulty with the notion of a class of discrimination problems, or the occurrence of a problem. Operational definition of w, the weight or probability of a cue, seems at first glance difficult, but since the theory makes it possible to evaluate θ for any problem, one can in principle measure the ratio of weights of any two sets of cues. Thus, the measurement of w does not offer a theoretical difficulty, however complex the experimental manipulations may become.

The empirical definition of a cue is roughly the following: k is a cue if and only if, when the subject is given appropriate training, then he learns to make differential responses based solely on k. Here appropriate training is the most efficient training program possible. Often we do not know what training program this is or how long training must be continued to get learning, with the result that empirical use of this definition is hindered. It does, however, give a fairly clear intuitive idea of the meaning of the term *cue*.

To define S^*, the set of problems, it is essential only to know what a cue is and to distinguish relevant from irrelevant cues. A cue is relevant in a particular problem if it can be used in that problem as the basis for consistently correct response. A cue is irrelevant if the problem is so designed that the cue cannot be used as the basis for consistently correct response.

The relation of occurrence, Q, of a problem, does not take into account whether the subject makes a correct or incorrect response. Given the concept of a problem, the notion of occurrence of a problem is clear since it corresponds to the usual experimental notion of a trial (especially in non-correction type training where one run through the apparatus or situation is considered a trial).

Another characteristic of this theory is the very strong assumption identifying θ with the relative weight of relevant cues. Without this assumption it would have been extremely difficult to evaluate the needed learning parameters, and experimental tests would have been complicated immeasurably. While one may be skeptical that such a convenient assumption would

be satisfied, it permits a coherent and powerful theory to be constructed. Having made a very useful simplifying assumption, the theorist can always retreat when the data demand it.

Finally, it may be noted that this theory in its present form does not account for that important class of experiments in which the relevant cues are reversed, i.e., where the formerly correct cue becomes incorrect, and the formerly incorrect cue becomes correct. Generalization to this field of data is needed to broaden the empirical base of the theory.

REFERENCES

1. Bush, R. R. and Mosteller, F. A model for stimulus generalization and discrimination. *Psychol. Rev.*, 1951, **58**, 413–423.
2. Eninger, M. U. Habit summation in a selective learning problem. *J. comp. and physiol. Psychol.*, 1952, **45**, 604–608.
3. Estes, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.
4. Estes, W. K. and Burke, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276–286.
5. Estes, W. K. and Straughan, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.
6. Lawrence, D. H. Acquired distinctiveness of cues: II. Selective association in a constant stimulus situation. *J. exp. Psychol.*, 1950, **40**, 175–188.
7. Lawrence, D. H. The transfer of a discrimination along a continuum. *J. comp. and physiol. Psychol.*, 1953, **45**, 511–516.
8. McKinsey, J. C. C., Sugar, A. C., and Suppes, P. Axiomatic foundations of classical particle mechanics. *J. rational Mechanics and Analysis*, 1953, **2**, 253–272.
9. Restle, F. A theory of discrimination learning. *Psychol. Rev.*, 1955, **62**, 11–19.
10. Wyckoff, L. B. The role of observing responses in discrimination learning. *Psychol. Rev.*, 1952, **59**, 431–442.

Manuscript received 6/9/54

Revised manuscript received 10/30/54

PSYCHOMETRIKA—VOL. 20, NO. 3
SEPTEMBER, 1955

THE OBJECTIVE DEFINITION OF SIMPLE STRUCTURE IN LINEAR FACTOR ANALYSIS*

LEDYARD R TUCKER
PRINCETON UNIVERSITY
AND
EDUCATIONAL TESTING SERVICE

Requirements for an objective definition of simple structure are investigated and a number of proposed objective criteria are evaluated. A distinction is drawn between exploratory factorial studies and confirmatory factorial studies, with the conclusion drawn that objective definition of simple structure depends on study design as well as on objective criteria. A proposed definition of simple structure is described in terms of linear constellations. This definition lacks only a statistical test to compare with possible chance results. A computational procedure is also described for searching for linear constellations. This procedure is very laborious and might best be accomplished on high-speed automatic computers. There is no guarantee that the procedure will find all linear constellations, but it probably would yield satisfactory results for well-designed studies.

The principle of simple structure, proposed by Thurstone as a solution to the problem of indeterminacy of position of axes in the factorial structure, has received wide support and use in factor analysis. There have been, however, a variety of criticisms including (1) a skepticism regarding whether this principle of simplicity did, in reality, adequately parallel nature, and (2) a feeling of disturbance at the subjectivity involved both in theory and in application. The first problem, that of the validity of the simple structure concept, may be settled only by experimental studies. It is the purpose of this paper to assist in solving the second problem, that of subjectivity, by attempting to develop a more objective and operational view of the simple structure concept.

Two major concepts of the nature of factors are used to justify the principle of simple structure. Thurstone's views might best be summarized by the following quotations: "In the interpretation of mind we assume that mental phenomena can be identified in terms of distinguishable functions, which do not all participate equally in everything that mind does ··· No

*This research was jointly supported by Princeton University, the Office of Naval Research under contract N6onr-270-20, and the National Science Foundation under grant NSF G-642. The author is especially indebted to Harold Gulliksen for his many exceedingly helpful comments and suggestions made during the course of this development. A debt of gratitude is also owed to Mrs. Gertrude Diederich, who performed many intricate calculations in the experiments on computing procedures. The author further wishes to express his appreciation to Frederic M. Lord and David R. Saunders, who read the manuscript and made a number of very useful suggestions.

assumption is made about the nature of these functions, whether they are native or acquired or whether they have a cortical locus." (**14**, p. 57.) "Just as we take for granted that the individual differences in visual acuity are not involved in pitch discrimination, so we assume that in intellectual tasks some mental or cortical functions are not involved in every task. This is the principle of 'simple structure' or 'simple configuration' in the underlying order for any given set of attributes." (**14**, p. 58.) Cattell (**3**) expresses a similar view. In contrast to the foregoing, Holzinger and Harman (**5**) express a variant view that factor analysis, as a branch of statistical analysis, conveys information in the original data with an aim of parsimony which should not be construed as a search for fundamental categories. Similarly, Vernon (**19**) takes a position that "· · · it should be clear that a factor is a construct which accounts for the objectively determined correlations between tests, in contrast to a faculty which is a hypothetical mental power." (**19**, p. 8.) Others have taken views on either of these two sides, with still others sticking to some middle ground. Since each of these views can be interpreted as yielding support for the desirability of simple structure, we believe that the definitions to follow could be derived from either view and will not distinguish between them. Some such view is necessary, however, as an initial step toward acceptance of the simple structure concept.

Relation Between Design of Factor Analysis Studies and Simple Structure

The factorial study of human behavior might best be conceived as a program of studies rather than in terms of isolated, separate studies. Each study should build upon the knowledge gained from previous studies and add further to the verified fund of knowledge. Early studies in some domain, or class of behavior, will be more exploratory in nature and be made with less perfected batteries of measures. As knowledge increases concerning the interrelations of the various behaviors in such a domain, it should be possible to construct more satisfactory batteries for factorial analysis. Confirmatory studies should aid in firmly establishing the factorial structure.

In exploratory studies a fully determined simple structure solution should not be expected and rotation of axes will probably be continued on subjective bases. There may well be an attempt to maximize the number of small, insignificant factor loadings; but some attention may also be given to interpretive possibilities. While some assistance may be obtained from analytic procedures, it seems inevitable that the rotation of axes for exploratory studies will remain an art. This paper does not attempt to present a method for rotation of axes to simple structure in exploratory studies. Rather, in contrast, the definitions and procedure to follow are to be conceived as applying primarily to the more perfected factorial studies.

A major premise of the present argument is that the objective definition

of a simple structure is dependent on both an adequate study design and on objective analytic criteria. Not all factorial studies may possess a simple structure, only those studies involving an appropriate battery of measures made on an appropriate sample of individuals. Some requirements set forth by the analytic criteria may be met only in the study design. It is desirable, however, that there be a maximum of freedom in the design of factorial studies so as to fit as many situations as possible. For example, an experimenter should be in a position to test objectively hypotheses concerning the relations of complex measures to factorially simpler ones. Thus, it is desirable that the analytic criteria permit complex variables and not limit the study design to factorially pure measures. The factorial simple structure needs to be unambiguously present, however, in the data. This is a function of the study design.

Requirements for Objective Definition of Simple Structure

Following is a proposed list of requirements for satisfactory objective criteria for simple structure. These requirements should be interpreted as applying to individual studies since invariance of factorial results over various changes in the population of individuals sampled and in the battery of measures is a matter for experimental verification. It will be noted, however, that small variations of factor loadings and projections from ideal values are permitted. These small variations from ideal might result either from random sampling error peculiar to the sample of individuals measured or from errors of approximation in the basic factorial model.

A second point to be noted is that a choice is made as to kind of projection employed relating test vectors to factors. In the case of correlated factors, orthogonal projections of test vectors on normals to hyperplanes are used. These orthogonal projections for a particular factor depend upon location of only the hyperplane for that factor and upon the test vectors. They are independent of the locations of all other hyperplanes. A further reason for this choice as to type of projection is that the square of this type of projection can be interpreted to represent the independent contribution of the factor to the variance of the variable.

a. Basic requirements

1. Emphasis is placed on a maximum concentration of vectors along hyperplanes, that is, on a maximum number of zero projections on normals to the hyperplanes, allowance being made for small variations in observed projections.
2. The vectors interpreted as being in each hyperplane span a space of $(r - 1)$ dimensions, allowance being made for small variations in observed projections, where r is the number of dimensions in the common-factor space.
3. Exactly as many simple structure factors are obtained as there are dimensions in the common-factor space.

b. Types of freedom explicitly permitted

4. Oblique factors are permitted.
5. A minority of highly complex measures whose vectors have projections on several, up to all, factors is permitted in the battery being analyzed.

c. Operational requirements

6. The choice as to which projections are to be interpreted as zero is made on objective grounds.
7. An objectively determined best fit to the data is involved.
8. The best fit is unbiased in the limiting sense that when the variance of projections interpreted as zero is small, the mean of these projections is near zero.
9. Statistical tests exist which indicate the plausibility of accepting any particular solution as a simple structure.
10. An automatic computational procedure is available for use with any particular study.

The first three, or basic, requirements relate as much to the study design as to the objective criteria for simple structure. Each factorial study for which there is to be an objectively defined simple structure should be so designed that the configuration of vectors satisfies these requirements. For the objective analytical criteria, on the other hand, these basic requirements form the essential framework. The first requirement parallels the concept of simple structure. The second requirement is necessary for the hyperplanes to be determinate. Consider, for example, a group of vectors for one hyperplane such that there was a two-dimensional space into which they only had small projections that could be interpreted as zero. The normal to the hyperplane could be located anywhere in this space and satisfy the first requirement. The location of the hyperplane would not be unique. In order for the location of the hyperplane to be definite it is necessary for the vectors in this hyperplane to have small projections into only one dimension, that of the normal to the hyperplane. The third requirement pertains most directly to the study design in the sense that there must be as many hyperplanes of vectors that satisfy the first two requirements as there are dimensions in the common-factor space. The study design should be such that the number of common factors extracted should be quite definite. When the third requirement is met by the study design, it is necessary, but probably not difficult, for the objective criteria to meet it also.

The types of freedom explicitly permitted in requirements four and five were selected because they touch on controversial, or possibly controversial, points. Factorial practice has been divided on the point of oblique versus orthogonal factors. It is the opinion of the author that in the present context maximum liberty should be permitted. Whenever it seems advisable, a restriction could be inserted to the effect that only orthogonal factors were permitted. This could be a function of the study being analyzed or of the opinion of the analyst. The case for complex measures has been previously

mentioned in this article. It is desirable for experimenters to be able to check in an objective fashion on hypotheses related to complex variables. Allowance for measures that have loadings on all factors is at variance with Thurstone's requirement (**14**, p. 335) that each row of the factorial matrix have at least one zero loading. In the opinion of the author this becomes an unnecessary restriction in case the basic requirements previously listed are met.

The last five, or operational, requirements relate to desirable aspects of objective criteria for simple structure. Requirement six could be met by the establishment of a range of projections, centering on zero, to be interpreted as negligible or zero projections. The limits for this range could be considered as generalized constants to be defined by the analyst on a priori grounds. A best fit of the data in some statistical sense as per requirement seven is certainly desirable. That this best fit should be unbiased, as per requirement eight, is also desirable. It is this requirement, however, that is likely to differentiate between an ideal objective criterion and various approximate ones. Requirements nine and ten are quite crucial, but at the same time may be the most difficult to satisfy. The statistical test of requirement nine is necessary for scientific acceptability, but it may be the last point to be solved for objective criteria for simple structure. The automatic computing procedures should be as economical as possible. It may be, however, that the computations for an ideal objective criterion will be so complex and extensive that such a criterion will be applied only to a few critical studies. Approximate criteria that involve simple computations might be adequate in many cases and would be highly desirable. Developments in high-speed computers, however, may influence the relative economies of the criteria.

Review of Previously Proposed Objective Criteria for Simple Structure

Turning next to an examination of proposed analytical definitions and procedures for a simple structure solution, Thurstone's equation for a simple structure will be considered first (**11**; **14**, pp. 354–356). Thurstone makes the interesting proposal that his equation 28 is the equation for a simple structure.

$$\prod_{p=1}^{r}\left[\sum_{m=1}^{r} a_m\lambda_{mp}\right] = 0, \tag{1}$$

where p indicates simple structure factors, r is the number of factors, m indicates reference factors, a_m is a coordinate of a point on reference factor m, and λ_{mp} is the direction cosine on reference factor m of simple structure factor p. This equation states, in essence, that the product of the projections for each vector separately on the normals to the hyperplanes should be zero. This could be accomplished by the existence of at least one zero projection

for each vector. A least squares function for determining a best fit of the equation to data is suggested in Thurstone's equation 32.

$$\sum_{j=1}^{n} \prod_{p=1}^{r} \left[\sum_{m=1} a_{jm}\lambda_{mp} \right]^2 = \phi, \tag{2}$$

where the notation is as above and j indicates tests. ϕ is to be minimized. No procedure is presented, however, for accomplishing this solution. Let us now consider this equation for simple structure in terms of our list of requirements for satisfactory objective criteria for simple structure. Zero projections are emphasized as per the first requirement. The second requirement is not necessarily satisfied, however, especially for batteries composed of very simple variables such that each vector might have a number of zero projections. Consider, for example, a battery composed of r groups of variables so that the vectors for each group form a separate cluster. In order for the vectors in each hyperplane to span a space of $(r - 1)$ dimensions, each hyperplane would have to pass through $(r - 1)$ of the clusters. This will, of necessity, result in each cluster being located in $(r - 1)$ of the hyperplanes. Thurstone's equation, however, may be satisfied by each cluster being located in only one hyperplane. Thus, each of the hyperplanes may be rotated so as to include only one cluster and not include vectors spanning an $(r - 1)$-dimensional space. In this way Thurstone's equation does not satisfy our second requirement.

Thurstone's equation of a simple structure does involve as many simple structure factors as there are dimensions in the common-factor space and, thus, satisfies our third requirement. Both requirements on types of freedom permitted are met, except for permitting vectors with projections on all factors. Oblique factors may be involved. The variables may be complex up to the point of, but excluding, variables with projections on all factors. Among the operational requirements category, only the seventh and eighth requirements seem to be met. Thurstone has suggested a least squares function for the best fit of the equation to the data and this function seems to be unbiased in the sense of requirement eight.

Carroll (**2**) has proposed an analytic procedure that seems closely related to Thurstone's equation of a simple structure. In his development Carroll proposes "· · · that a satisfactory criterion for an approximation to simple structure is the minimization of the sums of cross-products (across factors) of *squares* of factor loadings." He obtains for each vector the products of each pair of projections on normals to the hyperplanes, sums these products for each vector and then over all vectors. Our first requirement that zero loadings are emphasized is satisfied. By employing products by pairs of projections Carroll circumvents the difficulty of Thurstone's equation in reference to our second requirement. Carroll's criterion is not necessarily satisfied by just one zero projection for each vector; thus, the solution tends

toward having each hyperplane determined in $(r - 1)$ dimensions. The third requirement is also satisfied in that a complete set of factors are considered. In the area of types of freedom permitted, either oblique or orthogonal factors may be used. There is a relation, however, between the use of complex variables and obtaining an unbiased fit to the data (requirements five and eight). Following a presentation of illustrative applications of his criterion, Carroll points out the biasing effects of complex tests and concludes that "These considerations lead to the conclusion that the present criterion will probably work best for well-designed factor studies where there are a large number of factorially pure tests and a relatively small number of factorially complex tests." (**2**, p. 33.) Requirements seven and ten are satisfied in that Carroll presents an objectively determined best fit and a procedure for accomplishing it. The procedure is laborious, but might be programmed for electronic computers. Requirements six and nine are not satisfied, but might be so by further developments and definitions. We conclude that Carroll's proposal is highly promising as an approximate method. It does satisfy the basic requirements, and tends to do so also for the types of freedom permitted, but it has some undesirable properties in the operational requirements area such that we agree with Carroll that his method is to be considered as yielding an approximation to simple structure.

Saunders (**9**, **10**) has proposed a criterion for an approximation to simple structure involving the sum of fourth powers of factor loadings on orthogonal axes. Since it can be shown that Saunders' criterion is mathematically identical with Carroll's criterion discussed above when the orthogonal case is considered, we need not discuss Saunders' work extensively. In addition to an interestingly different and simpler computational procedure from that of Carroll, Saunders presents some comparisons of results from actual studies with results that were obtained from chance configurations of vectors. The results are quite promising.

Several other interesting recent publications involving closely related work to Carroll's development include articles by Ferguson (**4**), Neuhaus and Wrigley (**7**), and Pinzka and Saunders (**8**). Ferguson, starting from information theory, suggests using the sum of squares of products of factor loadings as a measure of parsimony, or lack of parsimony. Neuhaus and Wrigley in their quartimax method maximize the sum of the fourth powers of the factor loadings. A point of interest is their use of the Illiac (a high speed electronic computer). Pinzka and Saunders extended Saunders' solution to the oblique case. The discussion of the preceding two paragraphs applies directly to all three of these papers.

Thurstone in 1936 (**12**) proposed an analytic solution for simple structure involving a least squares solution of projections for a sub-group of variables for each hyperplane. The sub-group of variables was selected in terms of limiting sizes of projections on successive trials of an iterative procedure.

All of our requirements are met explicitly except two, three, and nine. The method used for selection of variables for the sub-groups allows the possibility that the essential dimensionality of the space spanned by the sub-group would be less than $(r - 1)$. By essential dimensionality we mean the number of dimensions in which some vectors for the sub-group have projections that would not be interpreted as zero (that is, less than the stated limits on size of projections used in selecting the variables). In our proposal, to be discussed later, an objective procedure is indicated which will circumvent our objection to this method of Thurstone. For Thurstone's method as he proposed it, we feel that failure to guarantee that the sub-group spanned an $(r - 1)$-dimensional space was a serious drawback which would make the method unacceptable. Requirement three could be met for each study by a succession of solutions, each involving location of a single hyperplane, until as many distinct hyperplanes were found as there were dimensions in the common factor space of the study. There is no guarantee, however, that all such hyperplanes could be found.

A variant of Thurstone's preceding procedure was presented by Horst (**6**), in which he maximized the ratio of the sum of squares of significant projections to the sum of squares of all projections for each hyperplane. This is mathematically equivalent to minimizing the ratio of the sum of squares of the non-significant projections to the sum of squares of all projections. Again the difference between significant and non-significant projections was made in practice on size of projection in successive trials. Comments on this method are identical with those on the preceding method.

Tucker (**16**, **17**) proposed non-analytical procedures making use of graphs and judgment of the analyst designed to insure that the sub-groups did span spaces of $(r - 1)$ dimensions. In that these procedures involve subjective judgments in the process of analysis they will not be evaluated here. Their importance here is that they did attempt to solve one of the more important problems in the determination of simple structure hyperplanes. It is possible by continually reducing the sub-group of variables to obtain a sub-group that will have non-significant projections in one direction. The problem is to guarantee that such a sub-group does have some significant projections in all directions orthogonal to the one for which the projections are non-significant.

Thurstone has recently proposed a still different type of objective procedure (**15**) in which a minimum weighted sum of projections is obtained. The weights are related to the projections by an arbitrary step function so as to emphasize near zero projections. This is a single-plane method in that one hyperplane is determined at a time. Although only projections on successive trial normals are used, the distinction between significant and non-significant projections is not a sharp break but rather a transition dependent on lower weights for projections of intermediate size. This will

increase the chance of involving variables spanning $(r - 1)$ dimensions in the determination of the hyperplane. In that the range of projections that receive finite weights is broad there is a chance that the solution could be biased in the sense used in our requirement eight. Vectors with significant projections on the normal could influence the location of the normal and thus produce a non-zero mean of non-significant projections even when the variance of the non-significant projections was low. We conclude that this latest objective method should be classified as an approximate procedure. It may be a very useful procedure, however, since the computations are quite simple and the results presented by Thurstone indicate good approximations to the desired results.

Definition of Simple Structure by Linear Constellations and Vector Masses

In the objective definition of simple structure proposed here a concept of linear constellations is employed. Consider the left half of Figure 1. This

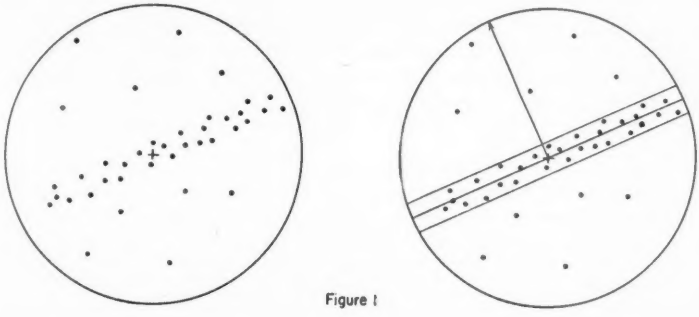

Figure 1

is a two-dimensional view of a factorial geometric model. Other dimensions are orthogonal to the plane of the figure. Each dot is the projection of the terminus of a vector representing a variable included in the battery being analyzed. It is postulated for Figure 1 that the battery of variables is such that the vectors might appear in a band such as is shown. If a direction is chosen orthogonal to this band, the vectors represented by the dots concentrated in this band will have small projections. In terms of a parametric explanation of the variances of the variables there will be a corresponding low dependence of these variables on a parameter corresponding to the direction orthogonal to the band. Such concentrations of vectors into linear spaces which include the origin may be termed linear constellations.

At the right of Figure 1, a line through the band of points and two bounding lines have been drawn to indicate the space of the linear constellation and the limits for projections outside this space. In general, linear constellations may be of any dimensionality less than that of the common-

factor space for the entire battery of variables. When the constellation contains only one dimension it would be called a cluster. This one dimension would represent a single parameter and could be interpreted. In case the linear constellation has as many dimensions less one as the common-factor space, the constellation may be designated by the one dimension orthogonal to the constellation. This normal can be used to indicate a parameter not involved in the constellation. The projections of the vectors on this normal will indicate the extent of dependence of the observed variables on this parameter. A simple structure is interpreted in the present context as a set of these linear constellations, the number of constellations in the set being equal to the dimensionality of the common-factor space.

The problem of defining simple structure is now transformed into that of explicitly defining linear constellation with dimensionality one less than the common-factor space. Let these constellations be termed linear constellations of dimensionality $(r - 1)$. Steps in the operational definition of such a constellation include the following:

1. Appended to and equally on both sides of any and every hyperplane in the common-factor space is a marginal space of some defined and limited width.
2. Any vector located entirely within a hyperplane and its marginal space shall be considered as contained in the hyperplane.
3. The number of vectors contained in a hyperplane shall be termed the vector mass of the hyperplane.
4. A maximum vector mass for a hyperplane occurs when rotation of the hyperplane in any direction results in a decrease in the vector mass before any subsequent increase in vector mass. (It is to be noted that with a finite number of vectors the location of the hyperplane for a maximum vector mass will not be unique. Small rotations of the plane may not result in a change in the vector mass.)
5. Those vectors contained in a hyperplane when the vector mass is maximum constitute a linear constellation of dimensionality $(r - 1)$ and the hyperplane will be termed the space of the linear constellation.

Definition of a simple structure adds the following step:

6. A simple structure is constituted by the hyperplanes for a set of r linear constellations of dimensionality $(r - 1)$.

A comparison of the foregoing definition with our requirements indicates that all requirements are met with the exception of number nine, relating to a statistical test, and number ten, concerning an automatic computational procedure. Emphasis is placed on a maximum concentration of vectors along the hyperplanes (requirement one). A maximum vector mass occurs only when the vectors contained in the hyperplane span a space of $(r - 1)$ dimensions, for otherwise a rotation would result in an increase in the vector mass (requirement two). In order to clarify this point, consider a group of vectors that are contained in a space of $(r - 2)$ dimensions and an appended space

of the defined radial width in the other two dimensions. In a three-dimensional factorial space such a group of vectors would form a cluster around a single direction. This group of vectors is contained in any hyperplane whose normal lies in the two-dimensional plane orthogonal to the given $(r - 2)$-dimensional space containing the group of vectors. Any hyperplane that contains just this group of vectors, therefore, may be rotated without loss of this group of vectors and may be made to contain one or more vectors not contained in the given $(r - 2)$-dimensional space. This step depends on the existence of vectors not contained in the $(r - 2)$-dimensional space, but such vectors must exist for the common-factor space to be of r dimensions. Thus, the vector mass of the hyperplane can be increased before any decrease occurs, and the original position of the hyperplane did not possess a maximum vector mass. This argument can be extended to vector groups contained in spaces of $(r - 3)$ or fewer dimensions. In consequence, a maximum vector mass occurs only when the vectors contained in the hyperplane are *not* contained also in a space of $(r - 2)$ or fewer dimensions; that is, the vectors contained in such a hyperplane must span a space of $(r - 1)$ dimensions.

The simple structure is defined in step six as being constituted by r hyperplanes, which is the dimensionality of the common-factor space (requirement three). No limitations are placed as to oblique or orthogonal factors or as to complexity of a minority of the tests (requirements four and five). A defined limit for projections of vectors to be contained in the hyperplane is indicated in our definitions one and two (requirement six). The linear constellations are objectively defined by maximum vector mass (requirement seven). This definition is unbiased since the marginal space of definition one is appended equally to both sides of the hyperplane (requirement eight).

It is hoped that one could derive a statistical test such as is indicated in requirement nine. Such a development would make a definite contribution to the field of factor analysis. At present, however, this requirement for a satisfactory criterion of simple structure has not been satisfied.

Computing Procedure for Linear Constellations

An automatic method for searching for linear constellations, as per requirement ten, has been developed and tried out. The labor of computations is quite great, but within bounds for automatic computing machinery. One trial has involved a run on an IBM Card Programmed Calculator. In addition a careful check has been made in detail on the feasibility of performing the computations on the IBM Type 701 Electronic Computer. This machine could perform the required computations on an automatic basis within feasible time, such as 10 minutes for 50 variables in 10 dimensions for each linear constellation.

It is of interest that the method finally adopted as feasible is a combination of two methods neither of which is feasible. The first of these methods

might be termed a direction survey method because it involved setting up a network of directions as trial normals to hyperplanes, computing the projections in each of these directions, and then determining the vector masses by counts of projections less in absolute value than some defined limit. Directions with maximum vector masses would be selected as normals to the spaces of linear constellations. Except in limited cases when the dimensionality of the space to be surveyed is small, the number of directions in even very rough networks becomes very great and this method is not feasible.

The second method involved various combinations of vectors as trial sub-groups. For each sub-group a direction could be determined such that the sum of squares of projections was a minimum. The largest sub-groups were selected which satisfied a condition that all members of each sub-group had projections less in absolute value than some limit on the direction with minimum sum of squares of projections for the sub-group. Because of the large number of combinations of variables to be considered for any study this method is not feasible.

Following is an outline of the computations for the combined method. These computations will be illustrated with material from a small study published by Thurstone (**11**, p. 167), who applied the centroid method to a table of intercorrelations published by Brigham (**1**, p. 275). There are fifteen tests and three dimensions. The reference factor matrix is given in Table 1. A series of successive approximation cycles are employed for each linear constellation. The outline of computations covers one of the cycles.

TABLE 1

Reference Factor Matrix, F*

Test No.	Arbitrary Orthogonal Axes I	II	III
10	.642	.443	-.150
2	.579	.499	-.090
5	.561	.449	-.041
3	.712	.228	.092
4	.633	.134	.061
1	.685	.159	.157
8	.529	-.144	.207
7	.559	-.146	.233
9	.546	-.222	.162
6	.585	-.293	.274
15	.475	-.112	-.132
14	.428	-.235	-.149
17	.619	-.303	-.194
11	.598	-.313	-.272
18	.436	-.084	-.099

*From: Thurstone, L. L. The vectors of mind. Chicago: University of Chicago Press, 1935.

TABLE 2

Computations of Two Smallest Principal Axes for an Initial Sub-group of Tests

Reference Factor Matrix, F_{g_1}

Test No.	I	II	III
5	.712	.228	.092
4	.633	.134	.061
1	.685	.159	.157

Matrix $P_{g_1} = F'_{g_1} F_{g_1}$

	I	II	III
I	1.3769	.3561	.2117
II	.3561	.0952	.0541
III	.2117	.0541	.0368

Matrix A_1 (Two Smallest Characteristic Vectors)

	C_1	C_2
I	-.288	.030
II	.871	.424
III	.398	-.905

Matrix $V_1 = F A_1$ (Projections of All Tests on Two Smallest Characteristic Vectors)

Test No.	C_1	C_2
10	.14	.34
2	.23	.31
5	.21	.24
3	.03	.03
4	-.04	.02
1	.00	-.05
8	-.20	-.23
7	-.20	-.26
9	-.29	-.22
6	-.31	-.35
15	-.29	.09
14	-.39	.05
17	-.52	.07
11	-.55	.13
18	-.24	.07

Step 1: List a matrix F_g for a selected sub-group of tests (see Table 2). For the first cycle the sub-group might be taken as those tests that have low correlations with some particular test. In experimental applications of this method, initial sub-groups were usually taken to contain approximately half of the tests in the battery. It was found that each of the linear constellations resulted from several different initial sub-groups. Enough different initial sub-groups were used for the study employed in the example to be able to find three distinct linear constellations. Two points of general concern are the recognition of duplicating results and being able to find all existing linear constellations. Any duplication can be readily detected by comparisons of the solutions and may be eliminated by discarding results from one or more initial sub-groups. The problem of selection of sub-groups so as to be able to find all linear constellations is much more difficult. After a number of constellations are found, a vector might be set orthogonal to them and tests selected that have low projections on this vector. Another possibility is to first employ a method such as Carroll's (**2**), or Saunders' (**9**, **10**) and to establish initial sub-groups of tests with low projections on each of the factors so determined.

The initial sub-group in the example contains tests 3, 4, and 1. For the second and subsequent cycles the sub-groups are given by the preceding cycle.

Step 2: Compute the matrix P_g (see Table 2).

$$P_g = F_g'F_g \,. \tag{3}$$

Step 3: Compute the two smallest characteristic vectors of P_g (see Table 2). These are the characteristic vectors corresponding to the two smallest characteristic roots of P_g . The smallest vector is C_1 and the next to smallest vector is C_2 . Each of these vectors is to be a unit vector (have sum of squares of entries equal to unity). The matrix containing these two vectors is labeled Λ in Table 2.

Step 4: Compute the matrix projections, V, of all the tests on the two smallest characteristic vectors (see Table 2).

$$V = F\Lambda. \tag{4}$$

Step 5: Survey the space of the two smallest characteristic vectors for the radial band of specified width which includes the largest number of test vectors. The concept involved is illustrated in Figure 2. A plot between projections of the tests on C_1 and C_2 is shown on the left. The dots for our trial sub-group of tests 3, 4, and 1 are located near the origin. Centered on C_1 and indicated by short lines outside the circle are eleven directions separated by 9°. The line with an arrow is pointing in the direction of $-36°$. Orthogonal to this trial normal is a line for the tentative linear subspace and two limit lines. The trial normal was also placed in each of the other

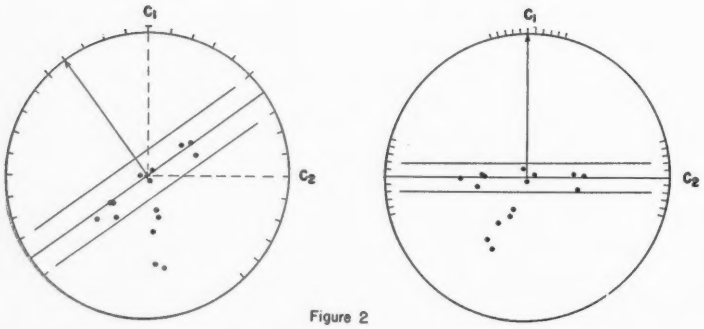

Figure 2

ten selected directions. The short lines inside the circle indicate the corresponding locations of the linear subspace. For each of the set of directions in this survey, a count was made of the number of points between the corresponding limit lines. For the direction in which the lines are drawn, ten of the dots lie in the space between the two limits. This would also be true for the trial normal placed at $-45°$. All other nine directions have fewer points in the space between the limits. The ten tests for the dots lying between the limit lines were selected for the next sub-group for the next cycle.

In practice, the plots in Figure 2 would not be made since the operations can be performed by computing steps illustrated in Table 3 and outlined below:

a. Define a transformation matrix U for the set of survey vectors to be employed. This matrix will contain the direction cosines of the survey vectors in terms of C_1 and C_2. Two such matrices are given in Table 5, a coarse survey set with 9° steps and a fine survey set with 3° steps. The coarse survey set was used in Table 3.

b. Find the projections of all tests on each of the survey vectors. These projections are contained in the matrix V_s of Table 3:

$$V_s = VU. \tag{5}$$

In this table, the test numbers for the sub-group of tests are double-starred.

c. Establish limits for projections to be considered as negligible and count the number of projections in each column of V_s within these limits. In the survey given in Table 3, limits of .15 and $-.15$ were used. All projections within these limits are starred, and the number of such projections in each column is given at the bottom of the table.

d. Choose the column of projections in V_s with the largest number of negligible projections. In case of a tie in this count between two columns, choose the column for the smallest angular deviation from C_1. In the example in Table 3, both columns $-45°$ and $-36°$ have counts of ten negligible projections. According to our rule to choose the column with the smallest angular deviation from C_1 we chose the $-36°$ column. A possible minor problem that may arise is when there is a tie between a column with a positive angular deviation from C_1 and a column with an equal negative angular deviation from C_1. In this case, an arbitrary decision might be made to choose the column with the positive angular deviation from C_1.

e. Select the tests with negligible projections on the chosen survey vector of step *d*

as the sub-group for the next cycle of computations. In the example of Table 3, the tests with starred projections in column $-36°$ were selected as the revised sub-group for the next cycle of computations.

It is anticipated that a rather coarse survey will be used during initial cycles for one of the factors, or linear constellations. When the sub-group of

TABLE 3

Coarse Survey for Initial Sub-group

Matrix $V_{s1} = V_1 U_c$

Test No.	Projections on Survey Vectors									
	$-45°$	$-36°$	$-27°$	$-18°$	$-9°$	$0°$	$9°$	$18°$	$27°$	$36°$	$45°$
10	-14*	-09*	-03*	08*	14*	19	24	28	31	34	
2	-06*	00*	07*	12*	18	23	28	31	34	37	38
5	-02*	03*	08*	13*	17	21	25	27	29	31	32
3**	00*	01*	01*	02*	02*	03*	04*	04*	04*	04*	04*
4**	-04*	-04*	-04*	-04*	-04*	-04*	-04*	-03*	-03*	-02*	-01*
1**	04*	03*	02*	01*	00*	-01*	-02*	-02*	-02*	-03*	-04*
8	02*	-03*	-07*	-12*	-16	-20	-23	-26	-28	-30	-31
7	04*	-01*	-06*	-11*	-16	-20	-24	-27	-30	-32	-33
9	-05*	-11*	-16	-21	-25	-29	-32	-34	-36	-36	-36
6	03*	-04*	-12*	-19	-25	-31	-36	-40	-43	-46	-47
15	-27	-29	-30	-30	-30	-29	-27	-25	-22	-18	-14*
14	-31	-35	-37	-39	-39	-39	-38	-35	-32	-29	-24
17	-42	-46	-49	-52	-53	-52	-50	-47	-43	-38	-32
11	-48	-52	-55	-56	-57	-55	-52	-48	-43	-37	-30
18	-22	-24	-25	-25	-25	-24	-23	-21	-18	-15	-12*
$\sum_n \lvert v \rvert <$	15	10	10	9	8	4	4	3	3	3	3	3	5

TABLE 4

Fine Survey for Revised Sub-group

Matrix $V_{s2} = V_2 U_f$

Test No.	Projections on Survey Vectors										
	$-15°$	$-12°$	$-9°$	$-6°$	$-3°$	$0°$	$3°$	$6°$	$9°$	$12°$	$15°$
10**	-17	-15	-14	-12	-10	-08*	-06*	-04*	-02*	00*	02*
2**	-09*	-07*	-05*	-03*	-01*	01*	03*	05*	07*	09*	11
5**	-06*	-04*	-02*	-01*	01*	03*	05*	07*	08*	10	11
3**	01*	01*	01*	01*	02*	02*	02*	03*	03*	03*	03*
4**	-03*	-03*	-03*	-03*	-03*	-03*	-03*	-03*	-03*	-03*	-03*
1**	07*	07*	06*	06*	06*	06*	06*	06*	05*	05*	05*
8**	09*	07*	06*	04*	02*	01*	00*	-02*	-04*	-05*	-07*
7**	10	09*	07*	06*	04*	02*	00*	-02*	-03*	-05*	-06*
9**	02*	00*	-01*	-03*	-05*	-07*	-09*	-11	-13	-14	-16
6**	11	09*	07*	04*	01*	-01*	-03*	-06*	-09*	-11	-13
15	-23	-24	-25	-26	-26	-27	-28	-28	-28	-29	-29
14	-26	-27	-28	-30	-31	-32	-33	-34	-35	-36	-36
17	-35	-36	-38	-40	-42	-43	-44	-46	-47	-48	-49
11	-42	-44	-46	-47	-49	-50	-51	-53	-53	-54	-55
18	-19	-20	-20	-21	-22	-22	-22	-23	-23	-23	-24
$\sum_n \lvert v \rvert <$	10	7	9	9	9	9	10	9	9	7	6

TABLE 5

Survey Matrices

Coarse Survey:

Matrix U_c

	Degrees Deviation from Smallest Characteristic Vector									
	-45	-36	-27	-18	-9	0	9	18	27	36	45
C1	.70711	.80902	.89101	.95106	.98769	1.00000	.98769	.95106	.89101	.80902	.70711
C2	-.70711	-.58779	-.45399	-.30902	-.15643	.00000	.15643	.30902	.45399	.58779	.70711

Fine Survey:

Matrix U_f

	Degrees Deviation from Smallest Characteristic Vector									
	-15	-12	-9	-6	-3	0	3	6	9	12	15
C1	.96593	.97815	.98769	.99452	.99863	1.00000	.99863	.99452	.98769	.97815	.96593
C2	-.25882	-.20791	-.15643	-.10453	-.05234	.00000	.05234	.10453	.15643	.20791	.25882

tests is not altered by a cycle of computations with a coarse survey, a finer survey may be employed. This finer survey would involve smaller angular steps for the survey vectors and narrower limits for negligible projections. Such a fine survey is illustrated at the right of Figure 2 for the illustrative problem and is given in Table 4. Three-degree steps were used in the second U matrix of Table 5, and limits of .10 and $-.10$ were used. In this case the sub-group for the cycle was composed of the tests indicated in Table 3 for the first cycle. These test numbers are double starred in Table 4. A series of fine surveys might be required before there is no change in the sub-group. When there is no change in the sub-group as illustrated in Table 4 (the 0° is the chosen column), the smallest characteristic vector, C_1 , is the normal to the desired hyperplane of the linear constellation, or factor.

This method has been tried on the illustrative example to determine three linear constellations by starting from different trial subgroups. These trial sub-groups were determined in this case as variables which had low correlations with selected variables. An alternative approach would be to apply one of the approximate solutions such as Carroll's (**2**) or Saunders' (**9**) and to pick variables with low projections on the resulting factors. In any case before the computations are initiated by the present method it is necessary to select a number of initial trial sub-groups and to define the two limits to be used in the coarse and fine surveys. Otherwise, the computations are completely automatic until a stable solution is obtained for each initial sub-group. At the end it will be necessary to compare the results from the several initial sub-groups and eliminate any duplications. In case the number of linear constellations discovered is less than the number of dimensions in the common-factor space, new initial sub-groups might be tried. Thus, this computing procedure is not so sure to find all of the linear constellations that are indicated in the definitions and may not satisfy our third requirement. It should yield satisfactory results, however, for those well-designed studies in which the vectors are concentrated along all hyperplanes.

REFERENCES

1. Brigham, C. C. A study of error. New York: College Entrance Examination Board, 1932.
2. Carroll, J. B. An analytic solution for approximating simple structure in factor analysis. *Psychometrika*, 1953, **18**, 23–38.
3. Cattell, R. B. Factor analysis. New York: Harper, 1952.
4. Ferguson, G. A. The concept of parsimony in factor analysis. *Psychometrika*, 1954, **19**, 281–290.
5. Holzinger, K. J. and Harman, H. H. Factor analysis. Chicago: Univ. Chicago Press, 1941.
6. Horst, P. A non-graphical method for transforming an arbitrary factor matrix into a simple structure factor matrix. *Psychometrika*, 1941, **6**, 79–100.
7. Neuhaus, J. O. and Wrigley, C. The quartimax method. *Brit. J. stat. Psych.*, 1954, **7**, 81–91.

8. Pinzka, C. and Saunders, D. R. Analytic rotation to simple structure, II: Extension to an oblique solution. Educational Testing Service Bulletin, RB-54-31, 1954. (Multilithed)
9. Saunders, D. R. An analytical method for rotation to orthogonal simple structure. *Amer. Psychologist*, 1953, **8**, 428. (Abstract)
10. Saunders, D. R. An analytical method for rotation to orthogonal simple structure. Educational Testing Service Research Bulletin RB-53-10 1953. (Multilithed)
11. Thurstone, L. L. The vectors of mind. Chicago: Univ. Chicago Press, 1935.
12. Thurstone, L. L. The bounding hyperplanes of a configuration of traits. *Psychometrika*, 1936, **1**, 61–68.
13. Thurstone, L. L. An experimental study of simple structure. *Psychometrika*, 1940, **5**, 153–168.
14. Thurstone, L. L. Multiple-factor analysis. Chicago: Univ. Chicago Press, 1947.
15. Thurstone, L. L. An analytical method for simple structure. *Psychometrika*, 1954, **19**, 173–182.
16. Tucker, L. R A rotational method based upon the mean principal axis of a sub-group of tests. *Psychol. Bull.*, 1940, **37**, 578. (Abstract)
17. Tucker, L. R A semi-analytical method of factor rotation to simple structure. *Psychometrika*, 1944, **9**, 43–68.
18. Tucker, L. R An objective determination of simple structure in factor analysis. *Amer. Psychologist*, 1953, **8**, 448. (Abstract)
19. Vernon, P. E. The structure of human abilities. London: Methuen & Co., Ltd., 1950. (New York: Wiley.)

Manuscript received 8/20/54

Revised manuscript received 11/15/54

PSYCHOMETRIKA—VOL. 20, NO. 3
SEPTEMBER, 1955

F-TEST BIAS FOR EXPERIMENTAL DESIGNS IN EDUCATIONAL RESEARCH

NEIL GOURLAY
UNIVERSITY OF BIRMINGHAM

Reference is made to Neyman's study of *F*-test bias for the randomized blocks and Latin square designs employed in agriculture, and some account is given of later statistical developments which sprang from his work—in particular, the classification of model-types and the technique of variance component analysis. It is claimed that there is a need to carry out an examination of *F*-test bias for experimental designs in education and psychology which will utilize the method and, where appropriate, the known results of this new branch of variance analysis. In the present paper, such an investigation is carried out for designs which may be regarded as derivatives of the agricultural randomized blocks design. In a paper to follow, a similar investigation will be carried out for experimental designs of the Latin square type.

I. *Introduction*

F-test bias may be said to exist for a given experimental situation if, when the null hypothesis is valid, frequent replication of the experiment provides a distribution of *F*-values which does not conform in some way (within the limits of sampling error) to the corresponding theoretical *F*-distribution. When bias exists, it is important for the investigator to know whether the *F*-test (the null hypothesis being valid) gives a larger or smaller proportion of significant *F*-ratios than is warranted by the theoretical distribution.

The possibility of *F*-test bias for certain experimental designs first became a topic of major statistical interest with Neyman's paper (**14**) in 1935. Neyman confined his inquiry to the randomized blocks and Latin square designs, which Fisher had developed; these designs had become the mainstay of agricultural experimentation. In both cases, he pointed out that "the conditions under which the application of the *z*-distribution is legitimate are not strictly satisfied" and went on to show that "in the case of the randomized blocks the position is somewhat more favorable to the *z*-test, while in the case of the Latin square this test seems to be biased, showing the tendency to discover differentiation when it does not exist."

Neyman's conclusions met at first with considerable opposition, but as Kendall (**10**, p. 214) points out, the controversy arose mainly from a failure to realize that Neyman was dealing with a different hypothesis from that usually tested. Thus, Fisher was concerned with the hypothesis that for each plot in the experimental field the treatments had the same effect. Neyman, on the other hand, stressed the possibility of interactions between plots and treatments and considered the more general hypothesis that the

mean effects of treatments over all plots involved in the experiment were the same. In 1937, Welch (**20**) clearly distinguished between the two hypotheses and went on to show that the z-distribution furnished an approximate test of the Fisher hypothesis both for the randomized blocks and Latin square designs. His findings did not, of course, invalidate in any way Neyman's analysis.

With regard to the validity of Neyman's analysis, it might be pointed out that while Neyman insists that the correction for fertility of a plot may vary from treatment to treatment, he regards the fertility corrections for blocks (in the case of randomized blocks) and for rows and columns (in the case of the Latin square) as being the same for all treatments. This assumption does not appear justifiable and, if not made, Neyman's results might be modified considerably.

Important as Neyman's investigation was for the randomized blocks and Latin square techniques, later years showed that his work was to exert a more widespread influence. Not only did he provide a method of analysis for detecting bias in special cases, but he aroused interest in the problem of bias generally. In addition, it was quickly realized that his method of analysis could also be employed to find estimates for the components of variance in any given experimental set-up. There arose a new branch of variance theory known generally as variance component analysis [cf. Crump (**5**)]. Further, the new approach made statisticians much more cognizant of the types of problem with which they had to deal: attention became focused on the types of mathematical model which they applied to different situations, and which formed the basis of their statistical analyses [cf. Eisenhart (**8**) and Crump (**5**)].

When we turn to consider the field of educational and psychological research it would appear that the randomized blocks and Latin square techniques were absorbed into this field without any noticeable recognition of the possible relevance of Neyman's findings. Recently, however, McNemar (**13**) has pointed out to users of the Latin square in psychology that they have ignored the fundamental assumption that all interactions are zero; and after stating, without giving or quoting any analysis in support, that failure to satisfy this assumption will lead to too many significant F's, he concludes that the Latin square technique is seldom appropriate and that "it is defensible only in those rare cases where one has sound a priori reasons for believing that the interactions are zero." Also it would be true to say that little reference, if any, has been made to the results of the other investigators who have continued the work that Neyman began. Two reasons might be offered in explanation. First, Neyman and many of the others were concerned with agricultural research and, consequently, were dealing with experimental situations which do not normally exist in education and

psychology. Secondly, many of the articles are of too recent origin for their results to appear to any great extent in the textbooks and research publications belonging to the latter field.

There is obviously a need to carry out an examination of F-test bias for experimental designs employed in education and psychology. Some such research has already been reported but it requires to be supplemented. The method and, where appropriate, the known results of variance component analysis should be utilized. In the present paper such an investigation is carried out for designs which may be regarded as derivatives of the agricultural randomized blocks design. In a second paper, the same will be done for those designs of which the agricultural Latin square is the prototype.

II. *Method*

For a valid (i.e., unbiased) F-test, the two variances involved in the F-ratio must be independent unbiased estimates—based on the stated numbers of degrees of freedom—of the same normal population variance. Test bias arises when the variances of the F-test fail to satisfy this set of conditions in one or more respects. It follows that, in order to detect bias, it is sufficient to examine the data—by simple inspection or by statistical analysis—for any failure to comply with the conditions of normality, homogeneity of variance, independence of estimates, etc.

It is not, however, sufficient to know that bias exists. An investigator also wants to know (a) the direction of the bias and (b) its magnitude; or, at least, he wants some indication of the answer to both these questions. For convenience, we will define an F-test to be positively or negatively biased, if, in the case where the null hypothesis being tested is correct, the test produces a larger or smaller proportion, respectively, of significant F-ratios than is warranted by the F-distribution.

In this paper considerable use will be made of Neyman's procedure (i.e., variance component analysis) as a method of detecting bias and of indicating its direction and magnitude. (The other and perhaps more common use of this form of analysis to obtain estimates of variance components will be involved only incidentally.) The method consists simply of taking the mathematical model which applies to the experimental situation and deriving analytically the expected values of the variances involved in the F-test. Then, in the case where the null hypothesis holds, the expected value of the "treatments" variance will be equal in magnitude to that of the "error" variance *if no bias is present.* When the two expected values are unequal, positive or negative bias is suggested according as the first variance is greater or less than the second. Also some measure of the magnitude of the bias is provided by the amount the ratio of the two expected values differs from unity.

For ease of exposition, we shall refer to this ratio as the B-ratio. Thus positive or negative bias is suggested by B-ratios greater or less than unity, respectively.

The method has several limitations:

(*i*) A B-ratio of unity is a necessary but not a sufficient condition for zero bias (the empirical F-distribution may have the same mean as the theoretical F-distribution but may differ from the latter in respect of standard deviation or any other moment). Consequently, it frequently happens that bias is present although the B-ratio is unity. Several instances of this appear in the present study.

(*ii*) As might be expected from (*i*), the value of the B-ratio is by no means always a certain criterion of the direction of bias; but it is probably true to say that it gives a correct indication of bias direction for most types of analyses where its value is other than unity.

(*iii*) In the same way, the deviation of the B-ratio from unity is only a very rough indication of the magnitude of bias. Obviously account must also be taken of the numbers of degrees of freedom involved in the F-test. A very tentative procedure for doing this will be suggested later.

Where Neyman's procedure is inadequate, other methods of bias analysis are required, and if these, because of the mathematical difficulty, are not easy to devise, empirical methods must be adopted. No such empirical studies are attempted in this paper, but reference is made to one or two studies of this type.

III. *Models*

All the models involved in this paper are linear models applying to a two-way classification. It may be helpful to the reader if, before proceeding to the main discussion, he is given an account of some of the models of this type to be found in the literature and is shown how the models of the present paper are related to them.

Eisenhart (**8**) distinguishes three types of models. In describing these we shall follow Crump (**5**) and adopt a broader interpretation than that chosen by Eisenhart.

Model I (*Fixed variate model*)

This may be written

$$X_{rst} = \mu + A_r + B_s + I_{rs} + \epsilon_{rst} \qquad \left\{\begin{matrix} r = 1, \cdots, p \\ s = 1, \cdots, q \\ t = 1, \cdots, n \end{matrix}\right\}, \tag{1}$$

where X_{rst} represents the tth observation in the subclass (r, s), μ is the general mean, A_r and B_s are the main effects for the corresponding column and row,

respectively, I_{rs} denotes the interaction effect for the rth column and sth row, and ϵ_{rst} is the random error for the observation.

The population of A's, B's and I's are all finite (of zero mean) and are exhausted in the given $p \times q$ classification, but the population of ϵ's is continuous with a normal probability distribution of variance σ_ϵ^2.

The expected values of the mean squares involved in the analysis of variance of the data are as follows:

	d. f.	Expected Value of Mean Square
Columns	$p - 1$	$\sigma_\epsilon^2 + qn \sum_r A_r^2/(p-1)$
Rows	$q - 1$	$\sigma_\epsilon^2 + pn \sum_s B_s^2/(q-1)$
Interaction	$(p-1)(q-1)$	$\sigma_\epsilon^2 + n \sum_r \sum_s I_{rs}^2/(p-1)(q-1)$
Residual	$pq(n-1)$	σ_ϵ^2

It will be seen that the null hypotheses (*i*) $A_r = 0$ $(r = 1, \cdots, p)$, (*ii*) $B_s = 0$ $(s = 1, \cdots, q)$, (*iii*) $I_{rs} = 0$ $(r = 1, \cdots, p;\ s = 1, \cdots, q)$ are tested by examining the significance of the F-ratios of *columns*, *rows*, and *interaction*, respectively, with respect to *residual*. Normally when *interaction* is significant, the investigator is not interested in making the test for *columns* and *rows* although there is no theoretical objection to his doing so. (Eisenhart actually restricts Model I to the case of zero interaction by making his second assumption of additivity).

Model II (*Random variate model*)

This may be written

$$X_{rst} = \mu + \alpha_r + \beta_s + \eta_{rs} + \epsilon_{rst} \qquad \left\{\begin{matrix} r = 1, \cdots, p \\ s = 1, \cdots, q \\ t = 1, \cdots, n \end{matrix}\right\}, \tag{2}$$

where the terms may be described as for the corresponding members of Model I but, in this case, the p α-values, q β-values and pq η-values are random samples from normal distributions of zero mean and of variance σ_α^2, σ_β^2, and σ_η^2, respectively.

The expected values for the mean squares in the variance analysis are:

	d. f.	Expected Value of Mean Square
Columns	$p - 1$	$\sigma_\epsilon^2 + n\sigma_\eta^2 + nq\,\sigma_\alpha^2$
Rows	$q - 1$	$\sigma_\epsilon^2 + n\sigma_\eta^2 + np\,\sigma_\beta^2$
Interaction	$(p-1)(q-1)$	$\sigma_\epsilon^2 + n\sigma_\eta^2$
Residual	$pq(n-1)$	σ_ϵ^2

The null hypothesis $\sigma_\eta^2 = 0$ is tested by testing *interaction* against *residual*. The hypotheses $\sigma_\alpha^2 = 0$, $\sigma_\beta^2 = 0$ are tested by testing *columns* and *rows* respectively against *interaction* or, where it is known *a priori* that interaction is zero, against *total* residual of $(pqn - p - q + 1)$ d. f. (It will be seen from the table that when $\sigma_\eta^2 = 0$, *columns* and *rows* can be tested against either *interaction* or *residual*. The latter provides the more precise test but a further increase in precision is obtained if the sums of squares for *interaction* and *residual* are combined to form an estimate of σ_ϵ^2 based on $(pqn - p - q + 1)$ d. f. and the tests of *columns* and *rows* made against this *total* residual).

Mixed Model

This takes the form

$$X_{rst} = \mu + A_r + \beta_s + \eta_{rs} + \epsilon_{rst} \qquad \left\{\begin{matrix} r = 1, \cdots, p \\ s = 1, \cdots, q \\ t = 1, \cdots, n \end{matrix}\right\}, \tag{3}$$

where the population of A-values is finite (of size p) but the β- and η-values are random samples from infinite populations. The expected values of the mean squares now read:

	d. f.	Expected Value of Mean Square
Columns	$p - 1$	$\sigma_\epsilon^2 + n\sigma_\eta^2 + nq \sum_r A_r^2/(p-1)$
Rows	$q - 1$	$\sigma_\epsilon^2 + n\sigma_\eta^2 + np\,\sigma_\beta^2$
Interaction	$(p-1)(q-1)$	$\sigma_\epsilon^2 + n\sigma_\eta^2$
Residual	$pq(n-1)$	σ_ϵ^2

Tests of hypotheses are made as for Model II.

Useful as the above classification is, it fails to cover many of the cases which occur in practice. Thus, the models of Fisher and Neyman for the agricultural randomized blocks design belong to quite a distinct class. A more extensive classification has been proposed by Tukey [cf. Crump (**5**)].

In the present paper, the models studied may be regarded as modified versions of Eisenhart's Mixed Model: there is only one exception, which is a special case of Model I.

The basic equation for these modified versions may be written

$$X_{rst} = \mu + A_r + \beta_s + \eta_{rs} + \xi_{rs} + \epsilon_{rst} \qquad \left\{\begin{matrix} r = 1, \cdots, p \\ s = 1, \cdots, q \\ t = 1, \cdots, n_{rs} \end{matrix}\right\}. \tag{4}$$

The main difference between this and Eisenhart's Mixed Model is the additional random error term ξ_{rs} common to all observations in the subclass (r, s). As will be seen later, this term differs from the η-term in that the ξ-values are usually regarded as independent (uncorrelated) while the η-values may be correlated and, what is more, show heterogeneity of correlation (between columns).

The next section, dealing with the investigation of bias for these models, falls conveniently into three parts:

1. Equal numbers in subclasses, i.e., $n_{rs} = \text{const.} = n$, say.

2. Numbers in subclasses unequal but proportional, i.e., $n_{rs} = Na_rb_s$ where N is total number of cases sampled and $a_1, \cdots, a_p$ and $b_1, \cdots, b_q$ are the proportions of cases in columns and rows, respectively ($\sum_r a_r = 1 = \sum_s b_s$).

3. Numbers in subclasses unequal and disproportionate.

Types of bias common to all three cases are discussed in the first part.

IV. *Investigation and Results*

1. *Equal Numbers in Subclasses* ($n_{rs} = n$)

It will make the discussion more concrete and less theoretical if we speak in terms of a methods experiment replicated in a random sample of schools. Lindquist (**12**) gives an excellent account of the experimental design and statistical analysis required for this type of experiment. The main F-test in the analysis is that of the *methods* variance against the *interaction* variance. The hypothesis tested is that the methods have the same mean effect over the *total* population of schools.

The interaction term of the analysis not only contains sampling error (measured by the variance *within classes*) but it may, and usually does, contain two other elements:

(*i*) real interaction between methods and schools;

(*ii*) group errors, i.e., errors which apply to the experimental groups as wholes and which are produced by factors other than method and school differences, e.g., teacher differences.

It will be seen that the model for this type of design is a version of the special mixed model mentioned at the end of the last section, namely,

$$X_{rst} = \mu + A_r + \beta_s + \eta_{rs} + \xi_{rs} + \epsilon_{rst} \qquad \left\{\begin{matrix} r = 1, \cdots, p \\ s = 1, \cdots, q \\ t = 1, \cdots, n \end{matrix}\right\}, \qquad (5)$$

where μ is the general mean and the A, β, η, ξ and ϵ represent the effects due to methods, schools, interaction, group error, and sampling error, respectively.

As usual $\sum_{r=1}^{p} A_r = 0$. ξ_{rs} and ϵ_{rst} are random, the parent populations

being assumed to be normal, of zero mean and of variance σ_ξ^2 and σ_ϵ^2, respectively. β_s is usually defined so that $(\mu + \beta_s)$ is the mean for the sth school over the p methods (and the total population of ξ- and ϵ-values). But it might be more instructive if we here take $(\mu + \beta_s)$ to be the mean of the sth school over a population of methods which includes the p methods under consideration. The parent population of β-values will be assumed to be infinite and of variance σ_β^2 (the mean of course is zero).

Since some of the methods within the total population of methods will normally resemble one another more than they do the others, the interaction terms for these methods (assuming there is real interaction between methods and schools) will be more highly correlated with one another than with the other η-terms. We shall now assume that for our p methods (and the total population of schools) the η-terms are equally correlated, with correlation ρ; we shall also assume that they are normally distributed with the same variance σ_η^2 for each method. The population mean will in each case be zero.

With this definition of our model the expected values of the mean squares for the analysis of variance are as in Table 1. For the benefit of the reader

TABLE 1

Variance	d. f.	Expected Value of Mean Square
Methods	$p - 1$	$\sigma_\epsilon^2 + n[\sigma_\eta^2(1 - \rho) + \sigma_\xi^2] + nq \sum_r A_r^2/(p - 1)$
Schools	$q - 1$	$\sigma_\epsilon^2 + n[\sigma_\eta^2(1 - \rho) + \sigma_\xi^2 + p\rho\sigma_\eta^2] + np\sigma_\beta^2$
Methods × Schools	$(p - 1)(q - 1)$	$\sigma_\epsilon^2 + n[\sigma_\eta^2(1 - \rho) + \sigma_\xi^2]$
Within Classes	$pq(n - 1)$	σ_ϵ^2

who is doubtful of the procedure for obtaining such a table, the derivation of the expected value of the mean square for methods is reproduced here.

The sum of squares between methods is given by

$$\sum_r qnM_r^2 - \frac{1}{pqn}\left(\sum_r qnM_r\right)^2 \qquad (r = 1, \cdots, p), \tag{6}$$

or more conveniently by

$$\frac{qn}{p} \sum_{k<l} (M_k - M_l)^2 \qquad (k, l = 1, \cdots, p), \tag{7}$$

where

$$M_k - M_l = \frac{1}{qn} \sum_s \sum_t X_{kst} - \frac{1}{qn} \sum_s \sum_t X_{lst} \qquad \begin{cases} s = 1, \cdots, q \\ t = 1, \cdots, n \end{cases}$$

$$= (A_k - A_l) + \frac{1}{q} \sum_s [(\eta_{ks} - \eta_{ls}) + (\xi_{ks} - \xi_{ls})] \tag{8}$$

$$+ \frac{1}{qn} \sum_s \sum_t (\epsilon_{kst} - \epsilon_{lst}).$$

Substituting in (7), squaring out and taking the expected value of the resultant expression, we obtain

$$\frac{qn}{p} \sum_{k<l} \left\{ (A_k - A_l)^2 + \frac{2}{q} [\sigma_\eta^2(1 - \rho) + \sigma_\xi^2] + \frac{2}{qn} \sigma_\epsilon^2 \right\} \quad (k, l = 1, \cdots, p),$$

which reduces to

$$qn \sum_r A_r^2 + n(p - 1)[\sigma_\eta^2(1 - \rho) + \sigma_\xi^2] + (p - 1)\sigma_\epsilon^2. \tag{9}$$

The required result follows.

If, instead of the above definition of β_s, we define β_s to be such that $(\mu + \beta_s)$ is the mean of the sth school over the p methods only, then $\sum_r \eta_{rs} = 0$ (as in the case of Model I). It is easy to show that ρ will now have the value $-1/(p - 1)$. [If we substitute this value for ρ in Table 1, the expected value for the *schools* variance becomes $(\sigma_\epsilon^2 + n\sigma_\xi^2 + np\sigma_\beta^2)$, which does not contain σ_η^2—a result which is obvious from the definition of β_s now being assumed.] It might be argued that the correlation ρ is an artifact, since its value depends on the way the β-values are defined and ρ can thus be made to have almost any value we please. But the reader should note that correlations in the variance component analysis cannot be avoided when there is heterogeneity of correlation between methods [case (c) below].

We will now consider three possible sources of bias for the *methods v. interaction* F-test. [Bias arising from non-normality in the data will not be considered in the present paper. Much work has been done in this field and, while most of it has been concerned with the simpler applications of the analysis of variance and not with more complex analyses such as may occur in education, it is probably true to say that these findings have general application.] An application of Neyman's technique to the modified form of the basic model for each of the three cases is useless, since the B-ratio is found to be unity. The results are not reproduced here. It is possible, however, to make some fairly definite pronouncements on the bias involved in each case.

Case (a). Heterogeneity of variance within classes (from school to school)

As a result of an empirical study, Lindquist and Godard (**12**, pp. 139–144) concluded that this type of heterogeneity "will not seriously affect the validity of the test of significance of methods differences based on the ratio of the M and $M \times S$ variance." A corollary to this result is that heterogeneity of group errors from school to school will not seriously bias the F-test.

Case (b). Heterogeneity of variance within methods

This type of heterogeneity may arise in two ways: either (*i*) the variance within classes may vary from method to method; or (*ii*) the variance due to "real" interaction may vary from method to method.

It seems unlikely, as Lindquist remarks (**12**, p. 144), that the methods would produce sufficiently large differences in variability to disturb the *F*-test seriously. But, where this did happen, the following remarks about bias might be made:

(*i*) No bias results from this type of heterogeneity when only two methods are involved. This can easily be established analytically.

(*ii*) With more than two methods, the bias is likely to be positive. It is known that when a *t*-test is applied to two random groups of the same size, heterogeneity of variance causes the test to be positively biased (**7**, p. 170). There is no contradiction between this result and that stated in the previous paragraph. Heterogeneity of variance produces bias in the *t*-test when applied to random groups but not when applied to matched groups. The latter case corresponds to the replicated experiment with two methods.

It is very likely that the same holds for the *F*-test when applied to more than two heterogeneous groups; and, if so, it would also apply to the (M *v.* $M \times S$)-test (when more than two methods are involved). A consideration of special cases adds support to this conclusion.

A discussion of this type of bias for a similar situation in agricultural research is to be found in Cochran and Cox (**4**, pp. 396–398). A more general discussion of the problem is to be found in Cochran (**2**). As a possible method of dealing with heterogeneity of variance, Cochran suggests the separation of the methods comparisons into single comparisons and the computation of separate error terms for each. It is better, however, if such a solution is found to be unnecessary (involving as it does a considerable loss in degrees of freedom).

Case (*c*). *Heterogeneity of correlation between class means* (*within methods*)

The point to be noted here is that some methods may be more alike than others, and, consequently, their interaction effects (with schools) will be more closely related with one another than with those for the other methods—thus producing heterogeneity of correlation between the class means within methods.

The type of bias present can be easily demonstrated with fictitious data for a highly theoretical case. (The example which follows probably affords a better understanding of the way in which the bias operates than is to be gained by any lengthy analysis).

Consider an experiment involving two methods, A and B, and seven schools. Let the means of the experimental groups be as follows:

	Schools						
	1	2	3	4	5	6	7
Method *A*	39	55	45	47	46	40	51
Method *B*	38	40	46	40	38	43	42

Then, with equal numbers in the experimental groups, the *methods* and *interaction* components of the variance analysis read:

	d. f.	Sum of Squares	Variance	*F*-ratio
Methods	1	87.5	87.5	4.375
Methods × Schools	6	120	20	

The *F*-ratio is not significant (for 1 and 6 d. f., $F = 5.99$ at the 5 per cent level of significance).

Now suppose that a third method, *C*, had been incorporated in the experiment and let us take the extreme case of *C* being identical with *B*. Also, let us imagine, to present the argument in its simplest form, that in this experiment there is no sampling error and that the interaction term consists only of real interaction. Then the means for the school groups subjected to method *C* will be the same as those for the groups undergoing method *B*. An analysis of variance for the three methods will therefore still give the same value for the *F*-ratio; but now with 2 and 12 d. f., significance is obtained at the 5 per cent level ($F = 3.88$).

We might consider what would happen with further replication of method *B*. Thus, with four replications, significance can be obtained at the 1 per cent level ($F = 4.22$ for 4 and 24 d. f.). Obviously, with the given form of analysis, the replication process increases the number of degrees of freedom without producing any real increase in the precision of the comparison of the methods. With a separation of the methods comparisons, such as Cochran suggests (see above), the spurious effect can be avoided.

The fact that methods are never identical and that sampling and other errors are always present does, of course, considerably reduce the amount of bias of this type which can occur. It is very probable that in most practical cases it is not serious. The use of covariance analysis or any other technique which improves precision by reducing random error will, of course, increase the importance of real interaction and so the type of bias under discussion. Covariance, etc., will also increase the effect of bias resulting from heterogeneity of variance of "real" interaction [see case (b) above].

Before concluding this section, two matters may be mentioned which are *not* irrelevant to the above discussion:

(*i*) As several writers have pointed out [e.g., Lindquist (**12**, p. 98); Webb and Lemmon (**19**)], similarities between methods may also operate in an *F*-test to mask other significant methods differences present, [i.e., speaking more technically, such similarities reduce the power of the *F*-test, cf. Johnson (**9**)]. Diamond (**6**) contends that the effect is normally small. It will be seen that, in the case of replicated methods experiment, both masking and case (c) bias may be present; it will also be seen that they are in opposition to

each other. Which predominates would depend on the relative importance of real interaction and random error.

(*ii*). It is to be observed that the analysis of variance of repeated measurements for a group of individuals is similar in form to that for the replicated methods experiment. *Individuals* correspond to *schools* and the sets of measurements (or *trials*) correspond to *methods*. Also the main F-test is *trials v. interaction* (individuals $\times$ trials) corresponding to the (M *v.* $M \times S$)-test of the methods experiment.

It follows that a somewhat similar discussion of bias is involved. Case (a) does not arise but cases (b) and (c) are applicable.

It is very likely that the bias arising from heterogeneity of correlation between interaction effects can be more serious in the repeated measurements analysis than in the other. Since the sets of measurements must succeed each other in time, this is bound to result in greater correlation between sets of measurements coming close together than those further apart; the heterogeneity of correlation will increase the greater the intervals of time between the measurements. Lindquist (**11**) covers himself on this point when he states that his treatment of the analysis depends on the assumptions that all individual regression lines are linear and parallel and that deviations from individual regression are normally distributed and of equal variance for all subjects. He regards Alexander's tests (**1**) as superior in that they provide for the possibility of individual differences in regression. Certainly Alexander's method of analysis is able to reveal any heterogeneity of individual regression which may be present. But Lindquist does not make the obvious point that, as a result of this heterogeneity, Alexander can only apply his F-tests to study trend for the group he was considering and not for the larger population with which Lindquist was concerned. There would appear to be two alternatives: either (a) to apply Alexander's method and so make a study of trend for the group only; or (b) to apply the simpler method of Lindquist to obtain a generalized result with the knowledge that, in certain cases, the result may be seriously biased.

2. *Proportionate Numbers in Subclasses* ($n_{r.} = Na_r b_.$)

It is generally accepted that difficulties in the application of the analysis of variance arise only with disproportionate numbers in the subclasses (the *nonorthogonal* case) and that proportionate numbers involve no more than slight computational changes of the procedure for equal numbers per subclass. This point of view is quite legitimate in the case where the hypothesis being tested has reference only to the rows and columns—whatever they represent—involved in the experiment (Eisenhart's Model I falls into this category). But it is inaccurate in the type of experiment—common in education—where the object is a generalized result which applies to a larger population (i.e., where Eisenhart's Mixed Model or a similar model-type applies). Where the

interaction variance has other components of variance besides that due to the variance within subclasses, proportionate numbers in the subclasses will introduce bias into the F-test of *treatments* against *interaction*.

This type of bias would appear to have been discovered first by Smith (**15**), who gives the results of a variance component analysis of Eisenhart's Model II with proportionate numbers in the subclasses. Concerned as we are here with experimental designs common in education, it will be more instructive if we consider the results for the special mixed model which underlies the methods experiment replicated over a number of schools.

The model is

$$X_{rst} = \mu + A_r + \beta_s + \eta_{rs} + \xi_{rs} + \epsilon_{rst} \qquad \left\{\begin{matrix} r = 1, \cdots, p \\ s = 1, \cdots, q \\ t = 1, \cdots, n_{rs} \end{matrix}\right\}, \tag{10}$$

where the symbols have the same meaning as in subsection 1, but now, with proportionate numbers in the subclasses, n_{rs} can be written as Na_rb_s, where N is the total number of cases and the a's and b's represent the proportions of cases corresponding to columns (methods) and rows (schools), respectively, $(\sum_r a_r = 1 = \sum_s b_s)$.

Since we have already dealt with the problem of heterogeneity of variance and correlation in the previous subsection, we shall assume homogeneity of variance and correlation for the present version of our model. The results of a variance component analysis are then as shown in Table 2.

Applying the null hypothesis, namely,

$$A_k = A_l \qquad (k, l = 1, \cdots, p), \tag{11}$$

we obtain for the B-ratio of the (M *v.* $M \times S$)-test the expression

$$(q-1)\frac{N(1-\sum_r a_r^2)(\sum_s b_s^2)S^2 + (p-1)\sigma_\epsilon^2}{N(1-\sum_r a_r^2)(1-\sum_s b_s^2)S^2 + (p-1)(q-1)\sigma_\epsilon^2}, \tag{12}$$

where

$$S^2 = [\sigma_\eta^2(1-\rho) + \sigma_\xi^2]. \tag{13}$$

Subtracting the denominator from the numerator of this expression we obtain the quantity

$$\begin{aligned} &N(1-\sum_r a_r^2)S^2[(q-1)\sum_s b_s^2 - (1-\sum_s b_s^2)] \\ &\quad = N(1-\sum_r a_r^2)S^2[\sum_{u<v}(b_u - b_v)^2] \qquad (u, v = 1, \cdots, q), \end{aligned} \tag{14}$$

which is positive except for the case in which the b's are all equal (when it becomes zero).

TABLE 2

Variance	d. f.	Expected Value of Mean Square
Methods	$p-1$	$\sigma_\epsilon^2 + \frac{N}{p-1}(1 - \sum_r a_r^2)(\sum_s b_s^2)[\sigma_\eta^2(1-\rho) + \sigma_\xi^2]$ $+ \frac{N}{p-1}\sum_{k<l} a_k a_l (A_k - A_l)^2 \qquad k, l = 1, \cdots, p$
Schools	$q-1$	$\sigma_\epsilon^2 + \frac{N}{q-1}(1 - \sum_s b_s^2)\{(\sum_r a_r^2)[\sigma_\eta^2(1-\rho) + \sigma_\xi^2] + \rho\sigma_\eta^2\}$ $+ \frac{N}{q-1}(1 - \sum_s b_s^2)\sigma_\beta^2$
Methods × Schools	$(p-1)(q-1)$	$\sigma_\epsilon^2 + \frac{N}{(p-1)(q-1)}(1 - \sum_r a_r^2)(1 - \sum_s b_s^2)[\sigma_\eta^2(1-\rho) + \sigma_\xi^2]$
Within Classes	$N-pq$	σ_ϵ^2

An immediate conclusion to be drawn is that inequalities among the b proportions (i.e., the proportions for schools) introduce bias into the F-test. Also from the fact that the B-ratio is greater than unity, it is likely that the bias is positive (special cases confirm this).

Having discovered this bias, we must ask: how does it arise? It is obviously due to the fact that the use of unequal proportions of pupils results in unequal weighting of the η- and ξ-terms when, of course, they should receive the same weighting. Also, for the same reason, inequalities among the a's must also produce bias although no indication of this is given by a consideration of the B-ratio alone. (It will be seen in fact that this effect is equivalent to that of heterogeneity of variance within methods).

How serious may the bias be? It will first be noted that, unlike the bias discussed in case (c) of subsection 1, the type of bias with which we are concerned here involves both the η- and ξ-terms which, together, are seldom negligible relative to the sampling error term (their relative effect will normally be increased by the use of covariance or a similar technique). Therefore, with large inequalities, the bias may be far from negligible.

An indication of the magnitude of the bias for unequal b proportions is the amount the B-ratio exceeds unity. This quantity can always be estimated for any practical case. To illustrate we shall use the data given by Lindquist in one of his examples (**12**, p. 120 *et seq.*).

$$N = 440 \qquad p = 4 \qquad q = 5$$

$$a_1 = a_2 = a_3 = a_4 = \tfrac{1}{4}$$

$$b_1 = \frac{10}{110}, \qquad b_2 = \frac{30}{110}, \qquad b_3 = \frac{19}{110}, \qquad b_4 = \frac{35}{110}, \qquad b_5 = \frac{13}{110}$$

The analysis of variance reads:

	d. f.	Sum of Squares	Variance
Methods	3	988.6	329.5
Schools	4	1748.3	437.1
Methods × Schools	12	172.8	14.4
Within Classes	420	2981.5	7.1

The entries in the last column may be taken as estimates of the corresponding expressions in the last column of Table 1. Thus, by simple arithmetic we obtain the following estimates:

$$\sigma_\epsilon^2 = 7.1; \qquad S^2 = [\sigma_\eta^2(1 - \rho) + \sigma_\xi^2] = .352;$$

$$\sigma_\epsilon^2 + \frac{N}{p - 1}(1 - \sum_r a_r^2)(\sum_s b_s^2)S^2 = 16.6$$

$$B\text{-ratio} = \frac{16.6}{14.4} = 1.15.$$

Note that for the given values of the b's, the value of the B-ratio cannot exceed 1.3, the value of $(q - 1) \sum_s b_s^2/(1 - \sum_s b_s^2)$.

However, the deviation of the B-ratio from unity is not in itself a good measure of the magnitude of bias. Allowance must be made for the numbers of degrees of freedom involved in the F-test. The greater the numbers of degrees of freedom, the more important a given deviation becomes and vice versa.

It is now tentatively suggested that in all applications of the B-ratio technique, the magnitude of bias is best measured by the expression

$$\frac{|B - 1|}{F_{1\%} - F_{5\%}}, \tag{15}$$

where $F_{1\%}$ and $F_{5\%}$ represent the values of F at the 1 per cent and 5 per cent levels of significance for the given numbers of degrees of freedom. It might then be established empirically, for any given design or model, how great the value of this expression must be before the bias becomes serious.

For the type of model discussed in this subsection, the bias due to inequalities among the b proportions can to some extent be overcome by testing *methods* not against the *interaction* variance provided by the straightforward analysis of variance but against the estimate of the expected value of the methods variance on the null hypothesis (i.e., for the given numerical example, against 16.6 instead of 14.4). [For a fuller account see Smith (**15**) and Cochran (**3**).]

3. *Unequal (Disproportionate) Numbers in Subclasses*

The literature on exact procedures for analyzing data of this type is now considerable. Tsao's paper (**18**) is probably the most rigorous and comprehensive. However, these methods involve much more computational labor than is demanded by the normal variance analysis. Also they have always been concerned with the testing of particular hypotheses, i.e., hypotheses having reference only to the rows and columns (whatever they represent) of the data to be analyzed (Eisenhart's Model I); and they have not, as yet, dealt with general hypotheses, i.e., hypotheses concerning a larger population of rows or columns (Eisenhart's Model II or Mixed Model). Investigations in educational research have therefore favored approximate methods of dealing with unequal numbers in the subclasses—at least where the criteria of applicability were satisfied. By far the most popular among these methods is Snedecor's Method of Expected Proportionate Frequencies (**16, 17**). In this section we will apply the B-ratio technique to investigate the bias which the use of this technique entails.

There will be the two cases to consider: (a) where the hypothesis tested applies only to the rows and columns of the data (Eisenhart's Model I);

(b) where a general hypothesis is tested, applying to a total population of rows or columns (Eisenhart's Mixed Model, etc.).

Case (a)

In this type of analysis, as was stated earlier in the paper, the two main F-tests are *interaction v. within subclasses*, and, when this test is not significant, *columns* (or *rows*) *v. within subclasses*. [Tsao (**18**) deals with other possible tests.] Before we proceed to derive the corresponding B-ratios, it is to be noted:

(*i*) Since Snedecor's method employs proportionate frequencies, the interaction term will not contain any component due to main effects (the characteristic of orthogonality), i.e., the interaction term is independent of the values of the main effects. Also the variance for columns will be independent of the main effects for rows and vice versa.

(*ii*) In deriving the first of the two B-ratios, we assume interaction to be zero; and in the case of the second, we not only make this assumption but we also assume zero differences between the main effects involved in the F-test.

It follows from (*i*) and (*ii*) that no serious loss of generality will be incurred (and a considerable saving in algebraic labor will be gained) if we straightway assume that interaction and the differences between main effects are zero; i.e., each observation, apart from a constant which we will here take to be zero, will consist only of sampling error and may be represented by

$$\epsilon_{rst} \quad \left\{ \begin{array}{l} r = 1, \cdots, p \\ s = 1, \cdots, q \\ t = 1, \cdots, n_{rs} \end{array} \right\}, \tag{16}$$

where p denotes the number of columns, q denotes the number of rows, and n_{rs} denotes the number of observations in the subclass (r, s).

It will be assumed that the ϵ's in all subclasses may be regarded as random samples from an infinite population of ϵ's of zero mean and variance σ_ϵ^2 (the usual assumption of homogeneity of variance). Thus, σ_ϵ^2 is the E. V. of the variance *within subclasses*.

Also let

$$\begin{aligned} Na_r &= \sum_s n_{rs} \\ Nb_s &= \sum_r n_{rs} \qquad (r = 1, \cdots, p; s = 1, \cdots, q). \\ Nc_{rs} &= n_{rs} \end{aligned} \tag{17}$$

Now let us derive the E. V.'s of the different sums of squares contained in Snedecor's Method. The sum of squares *between subclasses* is given by

$$\sum_r \sum_s N a_r b_s \left(\frac{1}{n_{rs}} \sum_{t=1}^{n_{rs}} \epsilon_{rst}\right)^2 - N\left[\sum_r \sum_s a_r b_s \left(\frac{1}{n_{rs}} \sum_{t=1}^{n_{rs}} \epsilon_{rst}\right)\right]^2, \tag{18}$$

which has E. V.

$$\sum_r \sum_s N a_r b_s \frac{\sigma_\epsilon^2}{n_{rs}} - N \sum_r \sum_s a_r^2 b_s^2 \frac{\sigma_\epsilon^2}{n_{rs}} = \sigma_\epsilon^2 \sum_r \sum_s \frac{1}{c_{rs}} (a_r b_s - a_r^2 b_s^2). \tag{19}$$

The sum of squares *between columns* is given by

$$\sum_r N a_r \left(\sum_s \frac{b_s}{n_{rs}} \sum_{t=1}^{n_{rs}} \epsilon_{rst}\right)^2 - N\left[\sum_r \sum_s \frac{a_r b_s}{n_{rs}} \sum_{t=1}^{n_{rs}} \epsilon_{rst}\right]^2, \tag{20}$$

which has E. V.

$$\sum_r N a_r \sum_s b_s^2 \frac{\sigma_\epsilon^2}{n_{rs}} - N \sum_r \sum_s a_r^2 b_s^2 \frac{\sigma_\epsilon^2}{n_{rs}} = \sigma_\epsilon^2 \sum_r \sum_s \frac{a_r(1 - a_r) b_s^2}{c_{rs}}. \tag{21}$$

Similarly the E. V. of the sum of squares *between rows* is

$$\sigma_\epsilon^2 \sum_r \sum_s \frac{b_s(1 - b_s) a_r^2}{c_{rs}}. \tag{22}$$

By subtracting the sum of (21) and (22) from (19), we obtain the E. V. of the *interaction* sum of squares

$$\sigma_\epsilon^2 \sum_r \sum_s \frac{a_r b_s (1 - a_r)(1 - b_s)}{c_{rs}}. \tag{23}$$

It follows that the *B*-ratio for the *F*-test *interaction v. within subclasses* is

$$\frac{1}{(p - 1)(q - 1)} \sum_r \sum_s \frac{a_r b_s (1 - a_r)(1 - b_s)}{c_{rs}}. \tag{24}$$

What values will this expression normally have?

It will first be noted that, when $c_{rs} = a_r b_s$, the *B*-ratio is unity since

$$\sum_r \sum_s (1 - a_r)(1 - b_s) = (p - 1)(q - 1). \tag{25}$$

This is, of course, to be expected since we are then dealing with proportionate numbers in the subclasses.

When $c_{rs} \neq a_r b_s$, the *B*-ratio may be positive or negative but a limited empirical study would suggest that it is normally positive. This again is to be expected for two reasons: (*i*) On the average the expression $a_r b_s / c_{rs}$ is likely to be greater than unity, since for a given difference between c_{rs} and

$a_r b_s$, the expression will exceed unity by a greater amount when $c_{rs} < a_r b_s$ than it will be exceeded by unity when $c_{rs} < a_r b_s$. (*ii*) The B-ratio, in any particular case, may be regarded as a weighted mean of the pq values of the expression $a_r b_s / c_{rs}$ for that case; it, too, will tend to have a value greater than unity.

It is, therefore, suggested here that the use of Snedecor's Method will normally produce a positively-biased test of interaction. The amount of bias will be indicated by the deviation of the B-ratio from unity or, better, by the value of the expression suggested at the end of subsection 2.

The B-ratio for the *columns v. within subclasses* F-test is

$$\frac{1}{p-1} \sum_r \sum_s \frac{a_r b_s (1 - a_r) b_s}{c_{rs}}. \tag{26}$$

It will be apparent that exactly the same can be said about this ratio as for the other.

Before we finish with case (a), it may be of interest to examine Tsao's modification of Snedecor's Method (**18**). Tsao "questions the validity of retaining the within variance derived from the original data while the other variances are derived from the adjusted data." To judge from the simplified case with which he deals at the end of his article, he would adjust the sum of squares *within subclasses* to the value

$$\sum_r \sum_s N a_r b_s \sum_t \left(\epsilon_{rst} - \frac{1}{n_{rs}} \sum_t \epsilon_{rst} \right)^2, \tag{27}$$

which will have E. V.

$$\sum_r \sum_s N a_r b_s \frac{(n_{rs} - 1)}{n_{rs}} \sigma_\epsilon^2 = \left(N - \sum_r \sum_s \frac{a_r b_s}{c_{rs}} \right) \sigma_\epsilon^2. \tag{28}$$

That is, the E. V. of his adjusted variance within subclasses is

$$\frac{1}{N - pq} \left(N - \sum_r \sum_s \frac{a_r b_s}{c_{rs}} \right) \sigma_\epsilon^2 \tag{29}$$

and not σ_ϵ^2 as for Snedecor's variance within subclasses.

If we are agreed that $a_r b_s / c_{rs}$ will on the average be greater than unity, it follows that the above E. V. will normally be less than σ_ϵ^2 . It would appear therefore that Tsao's correction will on the average increase the bias of Snedecor's Method.

Case (b)

It will clarify the discussion if we think of the columns as methods and the rows as schools. Our problem then is to investigate the bias of the (M *v.* $M \times S$)-test when Snedecor's approximate method is applied.

Obviously a part of the bias produced will be of the type discussed in

subsection 2 (provided, of course, there is either real interaction or group error). The rest of the bias will be of the type discussed in case (a) above. In order to study the importance of the latter type of bias for the (M *v.* $M \times S$)-test, let us take the special case where there is no real interaction and error consists only of sampling error. Then, from the analysis in case (a), it will be seen that the B-ratio for the (M *v.* $M \times S$)-test is

$$\frac{1}{p-1}\sum_r\sum_s \frac{a_r b_s(1-a_r)b_s}{c_{rs}} \Big/ \frac{1}{(p-1)(q-1)}\sum_r\sum_s \frac{a_r b_s(1-a_r)(1-b_s)}{c_{rs}}. \tag{30}$$

Since both numerator and denominator may be regarded as weighted means of the same pq values of $a_r b_s/c_{rs}$, it follows that the B-ratio will vary about unity, the degree of variation diminishing with the increase in number of the rows and columns. Therefore, as far as this type of bias is concerned, it is likely that Snedecor's Method will generally provide a more valid F-test for case (b) than for case (a).

The complete B-ratio for Snedecor's (M *v.* $M \times S$)-test, when both types of bias are involved, can be written down without further calculation (cf. previous subsection). It is

$$(q-1)\frac{N(1-\sum_r a_r^2)(\sum_s b_s^2)[\sigma_\eta^2(1-\rho)+\sigma_\xi^2]+\sigma_\epsilon^2\sum_r\sum_s\frac{a_r b_s^2}{c_{rs}}(1-a_r)}{N(1-\sum_r a_r^2)(1-\sum_s b_s^2)[\sigma_\eta^2(1-\rho)+\sigma_\xi^2]+\sigma_\epsilon^2\sum_r\sum_s\frac{a_r b_s}{c_{rs}}(1-a_r)(1-b_s)}. \tag{31}$$

An estimate of the value of this expression can easily be found for a given case.

In examining the bias increased by Snedecor's Method, no mention has been made of the χ^2 criterion for the applicability of the method. Snedecor established this criterion by empirical methods. Obviously the B-ratio, or rather some such expression as (15), could be established empirically as an alternative criterion. In dealing with the type of analysis discussed under case (b), it is possible that this alternative might prove superior.

V. *Summary of Results*

The basic model is

$$X_{rst} = \mu + A_r + \beta_s + \eta_{rs} + \xi_{rs} + \epsilon_{rst} \qquad \left\{\begin{matrix} r = 1, \cdots, p \\ s = 1, \cdots, q \\ t = 1, \cdots, n_{rs} \end{matrix}\right\}.$$

For a complete description of this type of model see p. 233. The main F-test is that of *columns* (the main effects of which are represented by the A-terms) against *interaction* (columns $\times$ rows).

1. *Equal Numbers in Subclasses* ($n_{r_s} = n$)

Three possible sources of F-test bias were considered:

(a) *Heterogeneity of variance within subclasses (from row to row)*. There is no evidence of bias in this case. The same applies to heterogeneity of variance of the ξ-effects from row to row.

(b) *Heterogeneity of variance within columns*. No bias arises when only two columns are involved. When there are more than two columns, the bias is likely to be positive (for definition of positive and negative bias see p. 229).

(c) *Heterogeneity of correlation between β-effects (of one column with another)*. The bias in this case is positive. In the typical methods experiment replicated in a number of schools it is unlikely to be serious; but for an analysis of variance of repeated measurements the bias involved might be considerable.

2. *Proportionate Numbers in Subclasses* ($n_{r_s} = Na_r b_s$)

For the given type of model (also for Eisenhart's Model II and Mixed Model), proportionate numbers in the subclasses produce bias, again of a positive character. The amount of bias depends on the degree of inequality among the a and b proportions; also on the magnitude of the η- and ξ-variances relative to the ϵ-variance. Gross inequalities in the proportions are obviously to be avoided in setting up experiments. A formula, of general application, is suggested for measuring the magnitude of bias.

3. *Disproportionate Numbers in Subclasses*

In this case, F-test bias was studied for Snedecor's Method of Expected Proportionate Frequencies. Eisenhart's Model I was considered as well as the mixed model stated above.

For Eisenhart's Model I, bias, if present, will normally be positive. When Tsao's modification of Snedecor's method is applied, it would appear that the bias will on the average be increased.

In the case of the Mixed Model, part of the bias arises in the same way as for Model I, but it is likely that, in general, it will not have the same importance. The other part of the bias is of the same nature as that discussed in section 2.

It is suggested that, for the Mixed Model, the expression for measuring bias proposed at the end of section 2, might prove superior to the χ^2-criterion as a test of the applicability of Snedecor's Method.

REFERENCES

1. Alexander, H. W. The estimation of reliability when several trials are available. *Psychometrika*, 1947, **12**, 79-99.
2. Cochran, W. G. Some consequences when the assumptions for the analysis of variance are not satisfied. *Biometrics*, 1947, **3**, 22-38.
3. Cochran, W. G. Testing a linear relation among variances. *Biometrics*, 1951, **7**, 17-32.

4. Cochran, W. G. and Cox, G. M. Experimental designs. New York: Wiley, 1950.
5. Crump, S. L. The present status of variance component analysis. *Biometrics*, 1951, **7**, 1-16.
6. Diamond, S. Comment on "A qualification in the use of analysis of variance." *Psychol. Bull.*, 1952, **49**, 151-154.
7. Edwards, A. L. Experimental design in psychological research. New York: Rinehart, 1950.
8. Eisenhart, C. The assumptions underlying analysis of variance. *Biometrics*, 1947, **3**, 1-21.
9. Johnson, N. L. Alternative systems in the analysis of variance. *Biometrika*, 1948, **36**, 80-87.
10. Kendall, M. G. The advanced theory of statistics, II. London: Griffin, 1946.
11. Lindquist, E. F. Goodness of fit of trend curves and significance of trend differences. *Psychometrika*, 1947, **12**, 65-78.
12. Lindquist, E. F. Statistical analysis in educational research. New York: Houghton Mifflin, 1940.
13. McNemar, Q. On the use of Latin squares in psychology. *Psychol. Bull.*, 1951, **48**, 398-401.
14. Neyman, J. Statistical problems in agricultural experimentation. *Suppl. J. roy. statist. Soc.*, 1935, **2**, 107-154.
15. Smith, H. F. Analysis of variance with unequal but proportionate numbers of observations in the subclasses of a two-way classification. *Biometrics*, 1951, **7**, 70-74.
16. Snedecor, G. W. The method of expected numbers for tables of multiple classification with disproportionate subclass numbers. *J. Amer. statist. Assoc.*, 1934, **29**, 389-393.
17. Snedecor, G. W. Statistical methods. Ames, Iowa: Iowa State College Press, 1946.
18. Tsao, F. General solution of the analysis of variance and covariance in the case of unequal or disproportionate numbers of observations in the subclasses. *Psychometrika*, 1946, **11**, 107-128.
19. Webb, W. B. and Lemmon, V. W. A qualification in the use of analysis of variance. *Psychol. Bull.*, 1950, **47**, 130-136.
20. Welch, B. L. On the *z*-test in randomized blocks and Latin squares. *Biometrika*, 1937, **29**, 21-52.

Manuscript received 6/18/53

Revised manuscript received 8/19/54

PSYCHOMETRIKA—VOL. 20, NO. 3
SEPTEMBER, 1955

LEAST SQUARES ESTIMATES AND OPTIMAL CLASSIFICATION

HUBERT E. BROGDEN
PERSONNEL RESEARCH BRANCH
THE ADJUTANT GENERAL'S OFFICE
DEPARTMENT OF THE ARMY*

A simple algebraic development is given showing that criterion estimates derived by usual multiple regression procedures are optimal for personnel classification. It is also shown that, for any assignment of men to jobs, the sum of the multiple regression criterion estimates will equal the sum of the actual criterion scores.

In earlier papers (**1**, **2**), the author contended that estimates of job proficiency derived by least squares estimates will place men in jobs in the most efficient way possible with the given predictor battery available, and that the average estimated job proficiency obtained by the use of such least squares estimates will equal the average actual job proficiency of assigned personnel. This paper will seek to establish these two points in a more rigorous fashion.

Definition of Symbols

C_{ij} = the performance of individual i in job j.

$\hat{C}_{ij}$ = estimates of the C_{ij}, each derived by regression equations from the same battery of tests and the same universe of individuals. It is assumed that the zero- and higher-order regressions involving the tests and the C_{ij} are linear.†

$\bar{C}_{ij}$ = the average C_{ij} value for a subset of individuals having the same pattern of scores on the battery of tests.

X = an allocation matrix with elements, x_{ij}, taking on values of zero and one. The x_{ij} entries for any individual have a single entry of one, and the x_{ij} entries for job j have Q_j entries of one. The remaining entries are zeros. The arrangement of ones in X corresponds to the placement of men in jobs. The use of X to symbolize any possible allocation of men to jobs is convenient and facilitates algebraic manipulation. In computing an allocation sum (to be defined), the cross-products of C_{ij} and x_{ij} are summed. When x_{ij} is one, the corresponding C_{ij} is included in the sum; when x_{ij}

*The opinions expressed are those of the author and are not to be construed as reflecting official Department of the Army policy.

†In practice, the C_{ij} would obviously not be available for all individuals in each job. Regression equations applying to the same universe can be estimated through a series of validation studies with a separate study being necessary for each job. In actual use the $\hat{C}_{ij}$ could then be computed for each applicant in each job.

is zero the corresponding C_{ij} is excluded. Thus, X represents any arrangement of zeros and ones, good or poor, consistent with the limitations already imposed, except that such an arrangement must be based solely upon the scores on the battery of classification tests. In other words, X represents any allocation of men to jobs consistent with the conditions of the problem.

K_j = a set of constants, one for each job. The K_j's are assumed to have numerical values such that, with allocation of each individual to the job in which $(\hat{C}_{ij} + K_j)$ is highest, the number allocated to each job will correspond to the number specified by the quota for that job.

X' = a particular X, with the x'_{ij} for each individual taking on a value of one for the job in which $(\hat{C}_{ij} + K_j)$ is highest. X' otherwise conforms to limitations imposed on X.

$\sum_{i,j} C_{ij}x_{ij}$ = the allocation sum. From the definition of an allocation matrix, it is evident that the allocation sum is equivalent to a simple sum, across all individuals, of the C_{ij}'s for the job to which each is assigned by a given allocation matrix.

Q_j = the quota for job j.

The Proof

We seek to demonstrate that

$$\sum_{i,j} \hat{C}_{ij}x'_{ij} = \sum_{i,j} C_{ij}x'_{ij} \geq \sum_{i,j} C_{ij}x_{ij} .$$

Consider a subset of individuals having an identical pattern of scores on the battery of tests basic to the $\hat{C}_{ij}$. Since we have specified that x_{ij} and x'_{ij} are to be based solely upon the test scores, it follows that both will remain constant in summing across individuals within such a subset. Then, for such a subset

$$\sum_{i,j} (\hat{C}_{ij} + K_j)x_{ij} = \sum_j (\sum_i \hat{C}_{ij}x_{ij} + \sum_i K_j x_{ij}) \tag{1}$$

$$= \sum_j (x_{ij} \sum_i \hat{C}_{ij} + \sum_i K_j x_{ij}). \tag{2}$$

Similarly, it follows t[b] at

$$\sum_{i,j} (\hat{C}_{ij} + K_j)x'_{ij} = \sum_j (x'_{ij} \sum_i \hat{C}_{ij} + \sum_i K_j x'_{ij}). \tag{3}$$

As N approaches infinity, the number in the subset approaches infinity. Now the criterion means of subgroups with identical score patterns are the basic data for graphic plotting of zero- and higher-order regression lines. If the regression system is linear, points representing the criterion means will fall on or near the regression lines. As the number in the subgroups approaches infinity the difference between $\bar{C}_{ij}$, the criterion mean for the subgroup, and $\hat{C}_{ij}$, the predictor value derived from a linear regression equation, will approach zero. Consequently, it is also true that, for the subset, $\sum_i C_{ij}$, the sum of the criterion scores, approaches equality to $\sum_i \hat{C}_{ij}$.

The basis for the equivalence of $\hat{C}_{ij}$ and $\overline{C}_{ij}$ within a subset having an identical pattern of scores might also be stated as follows: It is a basic principle of least squares prediction that the mean is the point at which the sum of the squares of the deviations is minimal. $\overline{C}_{ij}$, hence, is the best least squares estimate of the criterion scores of individuals with an identical pattern of test scores. If the regression system is linear, $\hat{C}_{ij}$ also provides the best least squares estimate. Hence, as N approaches infinity, the two must coincide.

From our definition of X', we know that, for such a subset

$$\sum_{i,j} (\hat{C}_{ij} + K_j)x'_{ij} \geq \sum_{i,j} (\hat{C}_{ij} + K_j)x_{ij} . \tag{4}$$

From equations 2 and 3,

$$\sum_j (x'_{ij} \sum_i \hat{C}_{ij} + \sum_i K_j x'_{ij}) \geq \sum_j (x_{ij} \sum_i \hat{C}_{ij} + \sum_i K_j x_{ij}). \tag{5}$$

Substituting $\sum_i C_{ij}$ for $\sum_i \hat{C}_{ij}$, we obtain

$$\sum_j (x'_{ij} \sum_i C_{ij} + \sum_i K_j x'_{ij}) \geq \sum_j (x_{ij} \sum_i C_{ij} + \sum_i K_j x_{ij}). \tag{6}$$

We may also write

$$\begin{aligned} \sum_j (\sum_i \hat{C}_{ij}x'_{ij} + \sum_i K_j x'_{ij}) &= \sum_j (\sum_i C_{ij}x'_{ij} + \sum_i K_j x'_{ij}) \\ &\geq \sum_j (\sum_i C_{ij}x_{ij} + \sum_i K_j x_{ij}). \end{aligned} \tag{7}$$

Since (7) holds for any subset, it holds in summing over all individuals.

In summing over individuals within any job, K_j is a constant and may be factored out. Both $\sum_i x'_{ij}$ and $\sum_i x_{ij}$ are, from the definition of X' and X, equal to Q_j . Hence, we have

$$\begin{aligned} \sum_j (\sum_i \hat{C}_{ij}x'_{ij} + K_j Q_j) &= \sum_j (\sum_i C_{ij}x'_{ij} + K_j Q_j) \\ &\geq \sum_j (\sum_i C_{ij}x'_{ij} + K_j Q_j) \end{aligned} \tag{8}$$

or

$$\sum_{i,j} \hat{C}_{ij}x'_{ij} + \sum_j K_j Q_j = \sum_{i,j} C_{ij}x'_{ij} + \sum_j K_j Q_j \geq \sum_{i,j} C_{ij}x_{ij} + \sum_j K_j Q_j \tag{9}$$

and, consequently,

$$\sum_{i,j} \hat{C}_{ij}x'_{ij} = \sum_{i,j} C_{ij}x'_{ij} \geq \sum_{i,j} C_{ij}x_{ij} . \tag{10}$$

We have, then, established two generalizations. First, we have shown that, as N approaches infinity, the predicted criteria for a set of jobs derived by the use of linear multiple regression equations yields, upon assignment of men to jobs, an allocation sum that is equal to or higher than that obtained

by any other assignment of individuals to jobs that is based on the test scores. Second, we have shown that, for any given assignment of men to jobs, the allocation sum obtained when regression estimates of the criterion are used becomes, as N approaches infinity, identical with that obtained when the criterion scores themselves are used.

REFERENCES

1. Brogden, H. E. An approach to the problem of differential prediction. *Psychometrika*, 1946, 11, 139–154.
2. Brogden, H. E. Increased efficiency of selection resulting from replacement of a single predictor with several differential predictors. *Educ. psychol. Meas.*, 1951, 11, 173–195.

Manuscript received 9/29/54

Revised manuscript received 11/15/54

PSYCHOMETRIKA—VOL. 20, NO. 3
SEPTEMBER, 1955

AN IMPROVED METHOD FOR TETRACHORIC *r*

W. L. JENKINS
LEHIGH UNIVERSITY

From the ratio of the cross-products of a fourfold table, with the application of two tabled corrections, tetrachoric *r*'s can be estimated with a mean discrepancy of less than .005 even when splits vary greatly from the medians. The necessary calculations can be handled by slide rule and the correction tables used without interpolation.

Davidoff and Goheen (1) have recently published a table for estimating tetrachoric *r*'s directly from the ratio of the cross-products of a fourfold table without correction. Unfortunately, the method gives accurate answers only when both distributions are split at approximately their medians. When the splits are not close to the medians, the obtained *r*'s are always biased in the positive direction. With some extreme splits, the positive bias amounts to .10, .15, or more.

However, it is possible to *correct* the obtained tetrachoric *r*'s by a method which is described and explained below.

Method and Example

1. Letter the fourfold table so that a is smaller than d and ad is greater than bc.

(*c*) 43	(*d*) 612
(*a*) 32	(*b*) 39

2. Compute the cross-products ratio ad/bc.

$$(32 \times 612)/(43 \times 39) = 11.68$$

From Table 1 find the uncorrected tetrachoric r for the nearest value of the cross-products ratio.

For 11.60, uncorrected $r = .756$.

3. Compute the two marginal splits $(a + b)$/total and $(a + c)$/total.

$$\frac{32 + 30}{726} = .10; \qquad \frac{32 + 43}{726} = .10.$$

TABLE 1

Uncorrected r from ab cd

ad/bc	r	ad/bc	r	ad/bc	r	ad/bc	r	ad/bc	r	ad/bc	r	ad/bc	r	ad/bc	r	ad/bc	r	ad/bc	r
1.00	.000	1.50	.158	2.00	.267	2.50	.344	3.00	.408	4.25	.517	5.50	.590	8.00	.683	12.0	.762	24.5	.861
1.02	.007	1.52	.163	2.02	.270	2.52	.346	3.05	.414	4.30	.520	5.60	.595	8.10	.686	12.5	.769	25.0	.864
1.04	.014	1.54	.168	2.04	.274	2.54	.349	3.10	.419	4.35	.523	5.70	.600	8.20	.688	13.0	.776	26.0	.868
1.06	.021	1.56	.172	2.06	.278	2.56	.352	3.15	.425	4.40	.527	5.80	.604	8.30	.691	13.5	.782	27.0	.873
1.08	.028	1.58	.177	2.08	.281	2.58	.355	3.20	.430	4.45	.530	5.90	.609	8.40	.693	14.0	.788	28.0	.876
1.10	.035	1.60	.182	2.10	.285	2.60	.358	3.25	.435	4.50	.534	6.00	.613	8.50	.695	14.5	.794	29.0	.880
1.12	.042	1.62	.187	2.12	.288	2.62	.360	3.30	.440	4.55	.537	6.10	.617	8.60	.698	15.0	.799	30.0	.883
1.14	.049	1.64	.192	2.14	.291	2.64	.363	3.35	.444	4.60	.540	6.20	.621	8.70	.700	15.5	.804	31.0	.886
1.16	.056	1.66	.196	2.16	.294	2.66	.366	3.40	.449	4.65	.543	6.30	.625	8.80	.702	16.0	.808	32.0	.890
1.18	.063	1.68	.201	2.18	.298	2.68	.368	3.45	.454	4.70	.547	6.40	.629	8.90	.705	16.5	.812	33.0	.893
1.20	.070	1.70	.206	2.20	.301	2.70	.371	3.50	.458	4.75	.550	6.50	.633	9.00	.707	17.0	.817	34.0	.896
1.22	.077	1.72	.210	2.22	.304	2.72	.374	3.55	.463	4.80	.553	6.60	.637	9.20	.712	17.5	.821	35.0	.898
1.24	.084	1.74	.215	2.24	.307	2.74	.376	3.60	.467	4.85	.556	6.70	.641	9.40	.717	18.0	.825	36.0	.901
1.26	.090	1.76	.219	2.26	.310	2.76	.379	3.65	.471	4.90	.559	6.80	.645	9.60	.721	18.5	.829	37.0	.904
1.28	.096	1.78	.224	2.28	.312	2.78	.382	3.70	.475	4.95	.562	6.90	.648	9.80	.725	19.0	.832	38.0	.906
1.30	.102	1.80	.228	2.30	.315	2.80	.384	3.75	.479	5.00	.564	7.00	.651	10.00	.730	19.5	.835	39.0	.909
1.32	.108	1.82	.232	2.32	.318	2.82	.386	3.80	.483	5.05	.567	7.10	.655	10.20	.734	20.0	.838	40.0	.911
1.34	.114	1.84	.235	2.34	.321	2.84	.389	3.85	.487	5.10	.570	7.20	.658	10.40	.737	20.5	.840	42.0	.915
1.36	.121	1.86	.240	2.36	.324	2.86	.391	3.90	.492	5.15	.572	7.30	.661	10.60	.740	21.0	.843	44.0	.919
1.38	.127	1.88	.243	2.38	.327	2.88	.394	3.95	.496	5.20	.575	7.40	.664	10.80	.743	21.5	.846	46.0	.922
1.40	.133	1.90	.247	2.40	.330	2.90	.396	4.00	.500	5.25	.577	7.50	.668	11.00	.746	22.0	.849	48.0	.926
1.42	.139	1.92	.251	2.42	.333	2.92	.399	4.05	.503	5.30	.580	7.60	.671	11.20	.750	22.5	.852	50.0	.929
1.44	.144	1.94	.255	2.44	.335	2.94	.401	4.10	.507	5.35	.583	7.70	.674	11.40	.753	23.0	.854	55.0	.935
1.46	.140	1.96	.259	2.46	.338	2.96	.404	4.15	.510	5.40	.585	7.80	.677	11.60	.756	23.5	.856	60.0	.940
1.48	.153	1.98	.263	2.48	.341	2.98	.406	4.20	.514	5.45	.587	7.90	.680	11.80	.759	24.0	.859	70.0	.945

In Table 2 at the intersection of the two marginal splits find the base correction.

Base correction = .103.

4. From Table 3 select a multiplier as follows: (a) If the larger split is less than .40, find the multiplier at the intersection of the difference between

TABLE 2

Base Correction

Larger Split	Smaller Split																					
	10	11	12	13	14	15	16	17	18	19	20	22	24	26	28	30	32	34	36	38	40	42	44	46
80	225	217	210	204	197	190	184	178	173	168	163													
78	217	209	204	196	190	184	177	171	166	160	156	147												
76	208	201	195	189	182	176	169	163	158	152	148	139	132											
74	201	194	187	180	174	168	162	156	150	143	140	131	124	116										
72	195	187	180	173	166	160	154	148	142	137	132	123	116	109	101									
70	188	180	174	166	160	154	148	142	136	131	126	117	108	100	092	084								
68	182	174	168	160	154	148	142	136	130	124	120	109	100	091	084	076	070							
66	175	168	162	155	148	142	136	130	124	118	113	103	093	084	076	069	062	056						
64	169	162	155	148	141	134	128	122	116	111	105	096	086	077	070	063	056	050	045					
62	163	155	148	141	134	128	122	116	110	105	100	089	080	072	064	057	050	045	040	035				
60	157	150	142	135	129	122	116	110	104	099	094	083	074	066	058	052	045	040	035	030	025			
58	151	144	136	129	122	116	110	104	098	093	087	077	068	060	053	046	040	035	030	025	020	016		
56	145	137	130	122	116	110	104	098	092	086	081	072	063	055	048	041	036	030	025	020	016	013	010	
54	139	131	123	116	110	104	097	091	086	081	076	067	058	050	044	037	032	026	021	016	012	008	005	002
52	134	126	119	111	105	098	092	087	081	076	072	062	054	046	040	033	027	022	017	012	008	004	000	000
50	129	121	114	106	100	094	088	082	076	071	066	058	050	042	030	029	023	018	013	010	006	000	000	000
48	124	116	108	101	095	088	082	076	071	066	062	054	043	038	032	026	020	015	010	008	004	000	000	000
46	119	111	103	096	089	082	076	071	066	062	057	050	041	036	028	022	017	013	009	007	003	000	000	000
44	114	106	098	091	084	078	072	067	062	057	053	046	038	031	026	020	016	012	008	006	002	000	000	
42	110	102	094	086	080	073	068	062	058	053	049	042	035	029	023	018	014	012	008	006	001	000		
40	105	097	090	082	076	069	064	059	054	050	046	039	032	026	021	016	013	011	007	005	000			
38	101	093	086	078	072	066	060	055	051	047	043	036	030	024	020	016	012	011	007	004				
36	098	090	082	075	069	063	058	053	049	045	041	034	029	023	019	016	012	010	007					
34	095	087	078	071	066	060	056	051	047	043	039	033	026	023	018	015	011	010						
32	091	183	075	168	063	058	054	049	045	042	037	032	026	023	018	015	011							
30	088	080	072	065	061	056	052	048	043	040	036	031	026	022	018	015								
28	087	079	070	063	058	055	052	048	043	040	036	031	026	022	018									
26	086	078	069	063	058	055	051	048	043	040	036	031	026	022										
24	085	077	068	063	058	054	051	048	043	040	036	031	026											
22	085	077	068	063	058	054	051	048	043	040	036	031												
20	085	076	068	063	058	054	051	048	043	040	036													
18	086	076	068	063	058	054	051	048	043															
16	089	079	070	064	059	055	053																	
14	092	082	072	066	061																			
12	097	087	075																					
10	103																							

TABLE 3

Multipliers for Base Correction

When larger split is more than .40, use Smaller split										Uncorrected	When larger split is .40 or less, use Difference between splits										When larger split is more than .40, use Smaller split					
10	20	30	40	50		r	00	05	10	15	20	25	30		10	20	30	40	50	
96	95	94	94	96		96	104	102	98	92	86	81	76		72	68	68	64	62	52
94	93	92	92	94		94	104	101	96	90	84	79	73		68	66	64	61	58	50
92	91	90	90	92		92	103	99	94	88	82	77	71		65	62	60	58	54	48
90	89	88	88	90		90	101	97	92	86	80	74	68		62	59	56	54	51	46
88	87	86	86	88		88	99	95	89	83	77	72	65		59	56	53	51	48	44
86	85	84	84	86		86	97	92	87	81	75	69	62		57	53	50	48	44	42
84	83	82	82	84		84	94	89	83	79	72	66	60		54	50	47	45	41	40
82	81	80	80	82		82	91	87	81	76	70	64	57		51	47	43	42	38	38
80	79	78	78	80		80	89	84	78	73	67	61	55		48	44	40	39	35	36
78	77	76	76	78		78	86	81	76	71	65	58	52		46	41	38	36	32	34
76	75	74	74	76		76	83	78	73	68	62	55	49		43	39	35	33	29	32
74	73	72	72	74		74	80	75	70	65	59	53	46		40	36	32	30	26	30
72	71	70	70	72		72	76	72	67	62	56	51	44		37	33	30	27	23	28
70	69	68	68	70		70	72	68	64	59	53	48	41		35	31	27	24	20	26
68	67	66	66	68		68	68	64	60	56	50	45	38		32	28	25	21	17	24
66	65	64	64	66		66	64	60	57	53	47	42	35		29	26	22	19	15	22
64	63	62	62	64		64	60	57	54	49	44	39	33		27	24	20	16	12	20
62	61	60	60	62		62	55	52	50	45	40	35	28		24	21	18	14	10	18
60	59	58	58	60		60	50	47	45	41	36	32	25		21	19	16	12	8	16
58	57	56	56	58		58	45	42	40	37	32	28	22		19	16	13	10	6	14
56	55	54	54	56		56	40	38	35	33	28	24	19		17	14	11	8	4	12
54	53	52	52	54		54	35	33	30	28	24	20	15		13	12	9	6	2	10

the splits and the uncorrected tetrachoric r. (b) If the larger split is .40 or greater, find the multiplier at the intersection of the smaller split and the uncorrected tetrachoric r.

Since the larger split is less than .40 in this example, use the difference of zero and the uncorrected r of .756 to find the multiplier of .90.

5. Multiply the base correction by the multiplier to secure the final correction.

$$.103 \times .90 = .093$$

6. Subtract the final correction from the uncorrected r to secure the corrected tetrachoric r.

$$.756 - .093 = .663$$

Explanation

Tables 1, 2, and 3 are derived from Pearson's tables of normal correlation surfaces (**2**). For Table 1, cross-product ratios for median splits were computed for r's of .05, .10, .15, $\cdots$, .95, and a curve constructed relating r to the cross-product ratios. The figures given in Table 1 are scaled from this curve.

Securing Tables 2 and 3 required a number of replottings of the Pearson data. Pearson's tables are set up in 0.1σ steps; decimal steps of marginal proportions are needed. Accordingly, it was necessary to pick values that corresponded roughly to the desired marginal splits at various levels of r and obtain cross-product ratios. These were plotted and replotted until a family of curves was obtained that related the needed corrections to three variables: the two marginal splits and the uncorrected tetrachoric r.

Table 2 is scaled from the family of curves according to steps of the two marginal splits, but for a single value of uncorrected tetrachoric r (.70). Except for such inaccuracies as may be introduced through repeated replottings, these corrections are precise when the uncorrected tetrachoric r is .70.

To avoid having a book of such tables (one for each step of uncorrected tetrachoric r), it was necessary to resort to some approximations. When both splits are small (below .40) the correction depends chiefly on the difference between the splits and the uncorrected tetrachoric r. When either split is large (above .40), the size of the smaller split (rather than their difference) has the greater influence. Table 3 is set up accordingly, presenting multipliers to be applied to the base corrections of Table 2.

Empirical check

The adequacy of the method is shown by the results of an empirical check involving the recomputation of 500 r's taken from the Pearson tables. Table 4 shows at the top the discrepancies of the uncorrected r's (all positively biased) such as would be obtained if Table 1 were used without correction. At the bottom are shown the residual discrepancies after the corrections of Table 2 and Table 3 have been applied. Even without interpolation, 88 per cent of the residual discrepancies are less than .005. With interpolation this rises to 94 per cent.

TABLE 4

Empirical Check on the Adequacy of the Correction Method

Discrepancies BEFORE correction (all positive)

True r	.000 to .020	.021 to .040	.041 to .060	.061 to .080	.081 to .100	.101 to .120	.121 to .140	.141 to .160	.161 and up
.10	30	15	1						
.20	23	19	11	3					
.30	16	16	11	5	4	2			
.40	11	12	11	10	6	3	2		1
.50	10	12	9	4	9	5	2	2	3
.60	9	11	7	6	7	4	4	3	5
.70	9	12	11	7	7	3	2	1	4
.80	9	12	6	6	3	4	4	3	
.85	12	8	7	6	6				
.90	12	9	3						

Discrepancies AFTER correction

True r	Without interpolation		With interpolation	
	.005 or less	More than .005	.005 or less	More than .005
.10	55	1	56	0
.20	52	4	54	2
.30	49	5	52	2
.40	51	5	51	5
.50	49	7	52	4
.60	50	6	53	3
.70	48	8	54	2
.80	42	5	42	5
.85	28	11	34	5
.90	19	5	19	5

REFERENCES

1. Davidoff, M. D. and Goheen, H. W. A table for the rapid determination of the tetrachoric correlation coefficient. *Psychometrika*, 1953, 18, 115-121
2. Pearson, K. (Ed.) Tables for statisticians and biometricians, Part II, London Biometric Laboratory, Univ. College, 1931.

Manuscript received 2/26/54

Revised manuscript received 8/16/54

BOOK REVIEWS

BENJAMIN FRUCHTER. *Introduction to Factor Analysis.* New York: Van Nostrand, 1954. pp. xii + 280. $5.00.

Several good books are already available in factor analysis. What claim can be made for another? Fruchter answers this in his preface. "These treatments have been found difficult by many otherwise competent students because of the mathematics and notation involved. It is hoped that this book will serve as an introduction to the subject and as a steppingstone to these more advanced texts."

The first four chapters provide a logical and mathematical introduction to factor analysis. Spearman's two-factor theory and its generalization to Holzinger's bi-factor method are discussed first. Cluster analysis is then considered as a means for understanding the logic of factor analysis. Next comes a chapter of "mathematics essential for factor analysis," including the basic matrix algebra operations and the geometry of rotation. This is followed by a chapter in which the basic equations of factor analysis are developed.

The next four chapters present the principal computational procedures. The diagonal and centroid methods are given in one chapter; the multiple-group and principal-axes methods in another; orthogonal rotation in a third; and oblique rotation in a fourth.

The final three chapters discuss (a) the interpretation of factors, (b) various applications of factor analysis, (c) some of the controversial issues in contemporary factor analysis. The book concludes with a useful bibliography of 700 titles covering principally the period from 1940 (the year of Dael Wolfle's review) to 1952.

Fruchter's statement of factor analysis differs in two main ways from the books already familiar to the readers of *Psychometrika.* First, his account is briefer and probably simpler than that of any of his predecessors; secondly, it has a better claim to be a textbook, less claim to be a personal statement.

Fruchter is undoubtedly right in saying that many otherwise competent students find factor analysis difficult because of the mathematics and notation. For many years to come, statements of factor analysis will be needed in which the approach is by means of the logic and calculations rather than by any rigorous mathematical development.

Fruchter stresses (a) the practical applications of factor analysis, and (b) the computations. Ten examples of the use of factor analysis are given in the chapter entitled "Applications in the Literature." These are of an interesting diversity, ranging from investigations of conditioned responses and rat maze learning to prepsychotic personality traits and Supreme Court voting records. *Q*- and *P*-technique are represented as well as *R*-technique. The chapter should be useful in reminding the psychological student that in studying factor analysis he must remain a psychologist. The computations in factor analysis are presented in detail in chapters 5 through 8. The various steps are itemized, and the instructions are for the most part clear and straightforward, so that the student who works diligently through the presentation should be able to calculate a factor analysis in a research of his own. The more experienced factor-analyst will probably be glad to have these step-by-step descriptions both for his own reference and for supplying to the student who seeks his aid.

A price is paid, naturally enough, for this emphasis upon learning by doing. For the most part the controversial issues of theory are eschewed. Key concepts are frequently introduced with so little discussion that the student may have trouble in seeing why the factor-analyst has adopted the particular procedure. For example, the account of communalities is brief and in my view very unsatisfying. The use of communalities is probably the factor-analytic procedure which has been most criticized by statisticians. The student whose knowledge is derived from this book will hardly be able to reply to any criticism.

The distinction between common and specific variance is initially made (p. 45) without any mathematical or logical reason being supplied for its adoption, and the brief discussions on pp. 46–47 and pp. 51–52 might well serve to confuse rather than clarify issues for the student. For one thing, Fruchter points out that communalities enable one to reproduce the correlations, and unities enable one to reproduce the original test scores; however, Fruchter provides no reason for preferring the former to the latter. For another, the information that specific variance is potentially common variance needs further development. As written at present, the distinction between the two types of variance is made to appear an entirely arbitrary one depending upon the particular selection of tests made by the investigator.

The discussion of orthogonal and oblique rotation is no more satisfactory. The distinction between simple axes and primary axes (i.e., factor structure and factor pattern) is deferred until the final chapter, which is a pot-pourri of theoretical issues set aside earlier. Yet it is doubtful whether the student will get any real understanding of the techniques of oblique rotation presented in an earlier chapter without knowledge of this distinction. Secondly, the controversy between those who favor orthogonal and those who favor oblique rotation is also held over to the final chapter. Even then the arguments for both sides are summarized very briefly, with Fruchter making no attempt to adjudicate upon the issues.

Let us next consider how this book differs from previous books. Each of these may have been referred to as a textbook, but invariably it has been a personal document as well. Thurstone's book, for instance, is primarily a statement of his original contributions and distinctive theories; little space is given to opposed views, except sometimes by way of rebuttal. Burt never allows his reader to forget that factor-analysts are by no means agreed in their theories and procedures and enters into logical and mathematical controversies with zest. Likewise, in Thompson, in Cattell, and in Holzinger and Harman space is found for personal contributions and points of view.

Perhaps "the battle of the schools" is ending in factor analysis. Fruchter's book has none of the intensity of debate characteristic of factor analysis in the thirties and the forties. Evidently many of the old disputes are settled. While the logic of factor analysis continues under discussion (as in Eysenck's and Hartley's recent articles), the degree of "reality" to be attributed to factors appears increasingly to be a metaphysical rather than a scientific issue.

For the already settled issues, Fruchter's avoidance of controversy is probably a strength. Factor analysis may have had overmuch of polemics in the past. It is in respect to the currently unsolved problems that Fruchter's approach seems to me a less happy one. The critical student who asks: "Is simple structure invariant?" or "Do the present tests of significance work?" or "How can we be sure that the rank of the matrix is reduced by the present means for estimating communalities?" does not get answers from Fruchter's text. Probably Fruchter cannot be expected to have answers to all of these, but at least they might have been indicated to be unsolved questions. The student who reads Fruchter alone can hardly know how many issues remain unsettled.

To summarize, Fruchter has set himself a limited objective. He has dealt very lightly with the mathematics and with the more theoretical issues of factor analysis. His emphasis is upon the calculations. Within these limits, Fruchter has done a good job. His survey is well balanced and impartial. For the student who needs to become familiar with the computations, the book will be very helpful. For the person who desires an understanding of factor analysis beyond that required for routine calculation, the book will not in itself be a sufficient guide.

University of Illinois *Charles Wrigley*

ANNE ANASTASI. *Psychological Testing.* New York: Macmillan, xiii + 682, 1954. $6.75.

The reviewer of a textbook serves essentially three functions. He attempts first to evaluate the soundness of the work from the point of view of accuracy, fundamental soundness, and good judgment in those areas where opinion rather than demonstrable knowledge is involved. Secondly, the reviewer must consider the book from the point of view of the audience for which it is intended and indicate whether he thinks it is suitable for the purpose stated. Finally, he must evaluate the book from the point of view of its original contribution to the total body of knowledge in the area covered.

Concerning the soundness of the book the reviewer finds remarkably little with which to take exception. The point of view presented is conservative and scholarly. With a few minor exceptions, the material seems to be accurate and precise. Where individual judgment and evaluation enter the picture, these judgments are on the whole conservative and, while pointing out weaknesses, tend for the most part to be favorable toward tests and test authors. While it is, perhaps, no truer in the field of testing than in other fields, it certainly can be said that the construction and publication of a test is an exercise in compromise between what is theoretically right and desirable on the one hand and what is practical and feasible in terms of time and expense on the other. While the author of this book is fully aware of the need for improvement and makes many suggestions as to how this may be brought about, there is nothing in the book to discourage the potential author from undertaking the construction of a new test or to discourage the test publisher from expanding his offerings.

The development of the book seems logical. Chapter II, "Principal Characteristics of Psychological Tests," lays the groundwork for what is to follow. Some psychologists might take exception to the definition of a psychological test as "essentially an objective and standardized measure of a sample of behavior" on the basis that this definition is too comprehensive, including as it does almost every possible variety of test. In the writer's opinion, achievement tests could be handled better as a separate category rather than as an aspect of psychological testing. Problems of reliability, validity, standardization, etc., are substantially different for achievement tests than for psychological tests in many instances. The American Psychological Association Test Standards Committee recognized this fact in leaving to the American Educational Research Association the production of a code for achievement tests.

Dr. Anastasi says that one can "consider all tests as behavior samples from which predictions regarding other behavior can be made. Different types of tests can then be characterized as variants of this basic pattern." She indicates further that one needs to be cautious in talking about measures of capacity, since capacity cannot be directly measured but can only be inferred from a measure of behavior. With this point of view, the writer of this review is in hearty agreement but he feels that the text has not gone far enough in indicating that many of the measures described are useful only if they are used to infer future behavior. This is certainly true of intelligence tests both of the general variety and the factor batteries and obviously true of prognostic and aptitude tests.

In the writer's opinion it may be considered one of the weaknesses of this text that insufficient attention is given to the basic problem of comparing such measures of capacity with subsequent measures of achievement. The problem of the criterion is discussed effectively but inadequate attention is paid to the problem of units in terms of which such pre- and post-measures can be compared. In fairness, it should be said that as much is done in this text as is generally done, perhaps more, in dealing with these problems. For example, a considerable section is devoted to expectancy charts, which is a noteworthy addition to what is ordinarily found in similar texts.

The sections of the book which deal with various types of tests are particularly

well done. The selection of tests used for illustrative purposes seems to be representative and sufficient information is given to provide the reader with a good notion of the various types of tests.

With regard to the evaluation of the book from the point of view of the *audience* for which it is intended, the writer cannot speak with such complete single-mindedness. Dr. Anastasi defines the audience as "the general student of psychology" and says further "Today, familiarity with tests is required not only by those who give or construct tests, but by the general psychologist as well." It is the considered judgment of this reviewer that this textbook cannot be read intelligently by psychology students taking a course in psychological testing without their having had at least an elementary course in statistics. Even with such a prerequisite, the book would appear to be more satisfactory for graduate rather than undergraduate classes and for students majoring in psychology rather than in education. This is contrary to the opinion stated by Dr. Anastasi in her preface where she says, "no previous knowledge of statistics is presupposed by the present text . . ." and ". . . for the benefit of students with no prior familiarity with statistics, however, all statistical concepts employed in the text have been explained and illustrated. Such statistical concepts have been introduced as they were needed and have been discussed within the appropriate context. Thus, they should appear more meaningful to the beginner than they would if segregated into a special 'statistical chapter.' " It appears to the writer that the section on reliability particularly and to some extent the sections on validity and norms will be completely incomprehensible to a person who has not had previous knowledge of basic statistics.

Dr. Anastasi indicates further that the book would be helpful to the practitioner in a number of fields, including the guidance counselor, school psychologist, psychometrist, personnel worker in business and industry and the clinical psychologist. With this point of view, the writer takes no exception. In fact, he would recommend the book as one which it would be very valuable for any practitioner to review and to have on his shelf for frequent reference purposes, especially if he has had a good grounding in statistics and elementary measurement.

As regards the third responsibility posed for the reviewer, namely, the evaluation of a book from the point of view of its original contribution, the writer of this review must conclude that there is little in this book that would appeal as being unique either in method, content, or emphasis. This can hardly be considered a serious indictment since originality is not the prime requisite of a good text. Original research, of course, is ordinarily reported in the professional journals or in professional papers, and in any generation the giants like Truman L. Kelley or Lewis M. Terman, whose books mark educational milestones, must necessarily be few in number.

Test Service and Advisement Center *Walter M. Durost*
Dunbarton, New Hampshire